ADIWES INTERNATIONAL SERIES

IN THE ENGINEERING SCIENCES

This book is in the

ADDISON-WESLEY SERIES IN
THE ENGINEERING SCIENCES
Dynamic Systems and Control Engineering

Consulting Editor:

J. Lowen Shearer

THEORY OF Automatic Control

M. A. AIZERMAN

TRANSLATED BY

RUTH FEINSTEIN

ENGLISH TRANSLATION EDITED BY

DR. E. A. FREEMAN

Sunderland Technical College
Sunderland, Co. Durham

1963

PERGAMON PRESS

OXFORD · LONDON · NEW YORK · PARIS

ADDISON-WESLEY PUBLISHING COMPANY, INC.

READING, MASSACHUSETTS · PALO ALTO · LONDON

U.S.A. Edition distributed by
ADDISON-WESLEY PUBLISHING COMPANY INC.
READING, MASSACHUSETTS, U.S.A.

Library of Congress Catalog Card Number 61-11523

This is a translation of the original Russian
Lektsiyi po teoriyi avtomaticheskogo regulirovaniya
published in 1958 by Fizmatgiz, Moscow

MADE IN GREAT BRITAIN

FOREWORD

THIS book is based on lectures which the author has given, over the last 10 years and longer, to scientific workers and engineers at a number of local scientific research institutes, design offices and factories in Moscow. His listeners were not specialists in the field of automatic control. They had had much experience of work in other spheres of technology and had mastered only the technology of control. Being convinced of the need to know the theory of automatic control, if only its essentials, in order to design and build controllers, they asked him to select material from the vast amount which has been accumulated in control theory, and to include in the lectures only what was most important for practical applications. In carrying out this request it is quite possible, of course, that the author has made mistakes in his selection of material. He has proceeded from his own experience, and the experience and advice of his listeners, but his opinions, habits and tastes could not but affect the content of the lectures. Only time and experience of practical applications enable us to separate basically what is essential from what is secondary.

The book is concerned with the theory, and not the technology, of control. The basic introduction to controllers in the first chapter is therefore long enough only for the following chapters to be understood.

The reader is expected to be familiar with the theory of the Laplace and Fourier transform. It is explained in Appendix 1 sufficiently for the reader who is not familiar with it to understand the book. Examples and some minor explanations have been put in small print, and may be omitted at a first reading.

A bibliography has been compiled separately for each chapter

and contains both source material and literature which may be useful to the reader for a deeper study of the theory of automatic control. The list includes mainly works published in the U.S.S.R. or translated into Russian. Of the non-Russian works not translated into Russian only direct source material has been included, since a complete list of the works which have been published in every other language connected with the problems discussed in this book would almost double its size.

The author is grateful to Ya. Z. Tsypkin, Yu. V. Dolgolenko, and A. Ya. Lerner for their comments on the manuscript, and to Yu. V. Krementulo, Yu. I. Ostrovskii and Ye. A. Andreyev for their help in preparing this book for press.

PREFACE TO THE SECOND EDITION

In the second edition, these lectures have been brought up-to-date and considerably amplified.

In Chapter I descriptions of discontinuous and oscillatory controllers as well as of extremal controllers, which have appeared in the last few years, have been included. The sections in Chapter III dealing with structural stability have been enlarged. A new section has been added to Chapter IV, concerned with statistical methods for analysing control systems when the disturbances are random. Chapter V, which is concerned with non-linear problems, has been almost doubled in size, mainly by the addition of an account of exact methods (where harmonics are not ignored) for investigating undamped oscillations.

The author takes the opportunity to thank his colleagues and readers in various towns of the U.S.S.R., as well as in the Federal German Republic, Japan and other countries, who sent their remarks and requests after the publication of the first edition of the lectures. The author has considered the larger part of them during the preparation of the book for the second edition.

CHAPTER I

GENERAL INTRODUCTION TO AUTOMATIC CONTROLLERS

1. General Considerations of Control Processes. Terminology

Automatic control is that process by which any quantity in a machine, mechanism or other technical equipment is maintained or altered according to given conditions without the direct participation of man, with the use of instruments called *automatic controllers* assembled specially for this purpose.

The fixed state of a machine or other equipment is usually disturbed by external actions, called *disturbances*. Whatever the nature of these disturbances, the *corrective action* of the controller must compensate for their disturbing action on the process. For example, the constant rate of revolution generated by a motor is disturbed by a change in the external load, and can be maintained only by action on the element which controls the supply of steam, fuel, etc.; the constant temperature in a room is disturbed by a change in the conditions of heat exchange, and must be compensated for by a change in the heating intensity in the room; the course of an aeroplane is disturbed by gusts of wind, by air pockets and other changes in the flight conditions, and must be maintained by means of action on the rudder, etc.

To keep a certain quantity constant, we could measure the disturbance and act on the machine in a way depending on these measurements. This method of process stabilization, called *automatic compensation*, is sometimes used, but when the sources of disturbance are very varied for one and the same machine, it is impractical. All the possible sources of disturbance can rarely be foreseen, moreover, they are often inside the machine itself (such as changes in lubrication

conditions, temperature and so on). Also, in many cases it is in general impossible to control the process using only automatic compensation.

To keep some quantity constant, instead of measuring the many different disturbances, we restrict ourselves more frequently to the measurement of the quantity which has to be controlled, and act on the machine in a way which depends on the deviation of this quantity from the desired value. The control then needs only one measurer however many different disturbances are present. Of course,

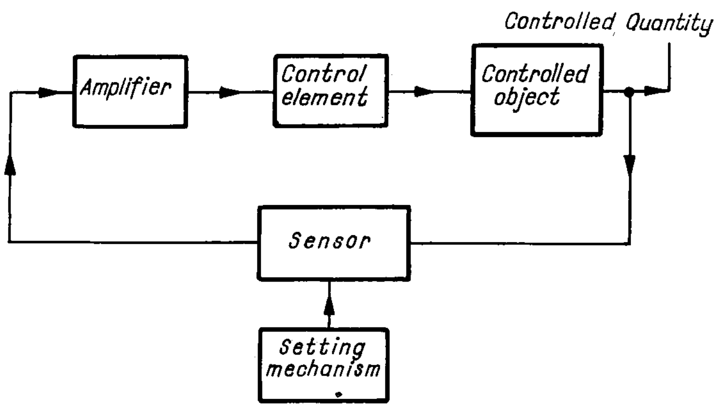


FIG. 1

the measured quantity cannot now be maintained absolutely exactly, since it is the deviation from the desired value which calls forth the corrective action on the machine. The fundamental scheme of a control system constructed on this principle is shown in Fig. 1.

The machine, mechanism or other equipment to which the controller is connected in order to put into effect the process of automatic control is called the *controlled object*, while the part of the controlled object which the controller acts upon is called the *control element.* The quantity which is subject to control is called the *control quantity* or the *controlled parameter.*

In order that the controller can cause the action on the control element it must contain a measuring instrument — the *sensor* of the controller — which measures the deviation of the controlled parameter from the desired value and a mechanism — the *setting mechan-*

ism — by means of which the desired value of the controlled quantity can be set up.

When the controller sensor can develop enough force and energy to move the control element with the necessary speed for any deviation of the controlled parameter from the desired value, it is connected directly to the control element. In this case, the controller is said to be a *controller of direct action*. Because of the exceptionally simple construction of direct action controllers they have until now been very common, but their use is restricted to objects which require little force to move their control elements.

In all other cases, the sensor is used only as a control instrument: a signal from the sensor controls an amplifier (hydraulic, electronic, metadyne and so on) which develops sufficient force and power to displace the control element by using an external supply of energy.

A controller which includes an amplifier of this kind is said to be a *controller of indirect action*.

The unit of an indirect action controller which is directly connected to the control element and sets it in motion is called the *regulating unit* of the controller. Various electric motors, hydraulic and pneumatic mechanisms, etc., can be used as regulating units.

We have already pointed out that no single controller can exactly maintain the desired value of the controlled quantity since, because of the principle of action itself, the controller only comes into operation after the process has been disturbed under the action of some disturbance or other. The changes of the controlled parameter brought about as the result of a disturbance and the controller action provoked by this disturbance are called the *control process*. The design of a controller must satisfy the strict technical requirements demanded by the nature of the control process. In addition, controllers usually possess a tendency to stir the process, to destroy its stability,* which has to be overcome during their design and adjustment. In order to guarantee stability of control and to facilitate the conditions under which the high demands made by the technical requirements on the control process can be satisfied, controllers are made more complicated by the introductions of various units specially intended for these aims. Such units include various internal feedbacks, with derivative, integral or forced action, with artificially introduced delays or breaks

* A definition of the term "stability" will be given in Chapter III.

in the action circuit, and so on. All units of this kind bear the name *stabilizing units*.

If the state of an object is disturbed by a change in the load on the controlled object and, further, the new load is maintained at a constant value (the residual load), then the controller may either restore the controlled parameter to the desired value independently of the magnitude of the disturbance, or it may set up a new value of the parameter which differs little (compared to the static error) from the old value, but which depends on the size of the residual disturbance. In the first case the controller is said to be *astatic*, and in the second static, with respect to this disturbance. The curve relating the value of the controlled parameter maintained by the controller to the residual load on the controlled object is called the *static characteristic of the control system*. In a number of cases astatic control must be used, but frequently static control with a small static error will be sufficient. As a rule static controllers prove to be more stable, and they ensure that the process is of high quality because they use simpler, and consequently also cheaper, methods of stabilization. Sometimes static control is essential for the correct operation of the objects as, for example, in the control of several machines working in parallel.

The account given above is directly concerned with controllers which are constructed to control a single variable. In industrial practice it is often necessary to control several variables in one object. If the variables are not connected to one another through the object, that is, if a deflection in any control element is caused by a deviation in one controlled variable only, and not by deviations in any other controlled variables, then each variable is controlled by its own controller. In this case the controllers assembled in the object act independently of one another.

It is often necessary to control objects in which the controlled variables are interdependent. A deflection in one control element for such objects causes a deviation in several of the controlled variables. If the object contains several independent controllers, then the operation of one of the controllers will cause a deviation in the other controlled variables, thus setting the other controllers into operation. Under these conditions, the controllers may interfere with one another.

For example, when controlling intermediate steam-selection turbines used for district heating plants, both a controller for the

angular velocity of the rotor and controllers for the steam pressure in the intermediate chambers are used. A pressure change in one of the chambers causes the pressure in the other chambers to change and also changes the angular velocity of the rotor. Conversely, a change in the load on the rotor changes not only its angular velocity but also the steam pressure.

It is often found advisable in such cases to connect the separate controllers in a single control system. In a connected system the control process of each controller controls not one but several or all of the control elements.

Such systems are said to be *multiple*. In multiple systems, the control process for any one quantity cannot be considered independently of the other variables, and the system as a whole must be considered as a single multiple complex. In a number of cases the connexions between the separate controllers can be chosen so that a deviation in one of the controlled quantities does not cause a deviation in the others, despite the mutual connexions through the object between the controlled quantities. They are compensated for by external couplings between the controllers. These systems are said to be *independent*.

In the following sections we describe automatic control units of basic types, giving simple examples, mainly of controllers of technological processes. The reader will find many more complex, and also more specialized, examples in specialist manuals and textbooks.

2. The Simplest Systems of Direct Control

Direct action controllers were the first controllers to become widely used.

Because of the exceptional simplicity of their construction and their operational reliability they have, until now, been widely used to control the angular velocity of revolving parts of machines, the level of liquid in various capacitances and the gas pressure in gas mains and gasometers.

Figure 2 gives a typical example of the layout of a centrifugal direct action controller. When the angular velocity of the machine shaft deviates from a given value (fixed by altering the tension in the spring) the centrifugal force of the weights changes and the equilibrium position of the coupler connected to the control element of the machine changes correspondingly.

Figure 3 gives another example of a centrifugal direct action controller. The weights in this controller are small balls held down by a spring between the plane and the conical plates. When the centri-

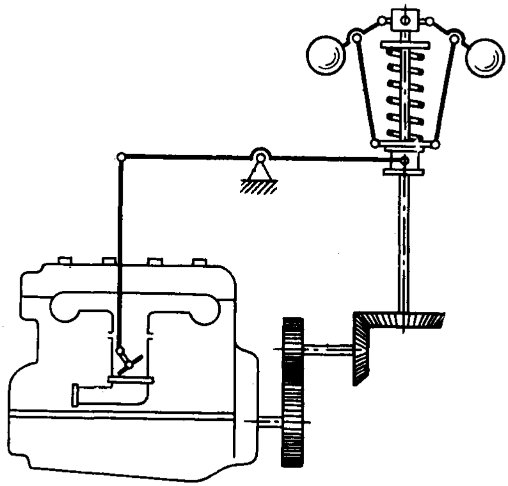


Fig. 2

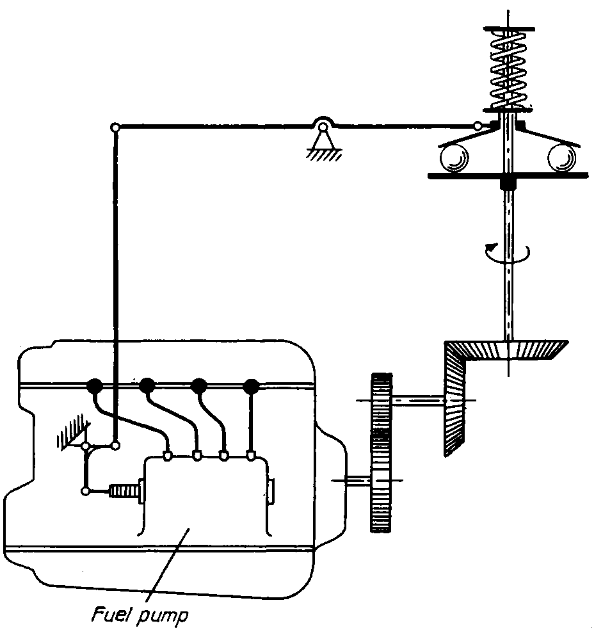


Fig. 3

fugal force changes, there is a change in the distance between the centre of the ball and the axis of revolution.

Unlike these so called conical controllers, plane controllers (Fig. 4) are such that the weights move in the plane of revolution.

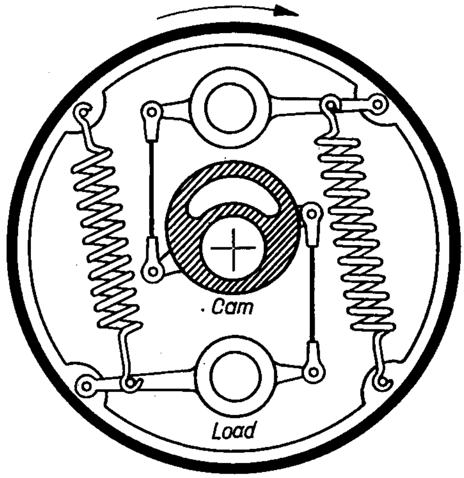


FIG. 4

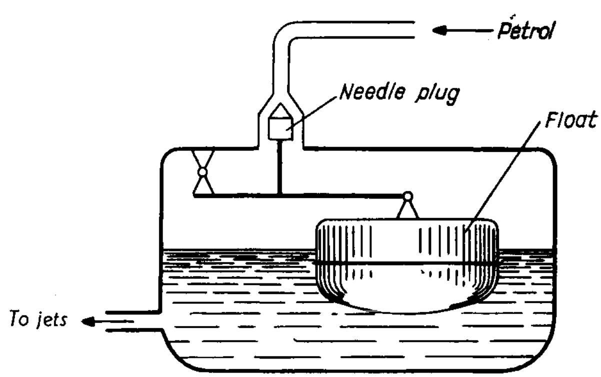


FIG. 5

As a result, the forces on the weights consist not only of the centrifugal force of inertia, which depends on the angular velocity, but also of the tangential force of inertia, which is proportional to the angular acceleration. The instantaneous position of the coupler and, of course, of the control element is determined not only by the deviation of the

controlled parameter (i.e. the angular velocity) from the desired value, but also by the size of its first time-derivative (i. e. the angular acceleration). The controller in Fig. 4 is an example of a direct action controller with derivative action.

Together with centrifugal controllers, float level direct action controllers are common. They are used very widely, from those in

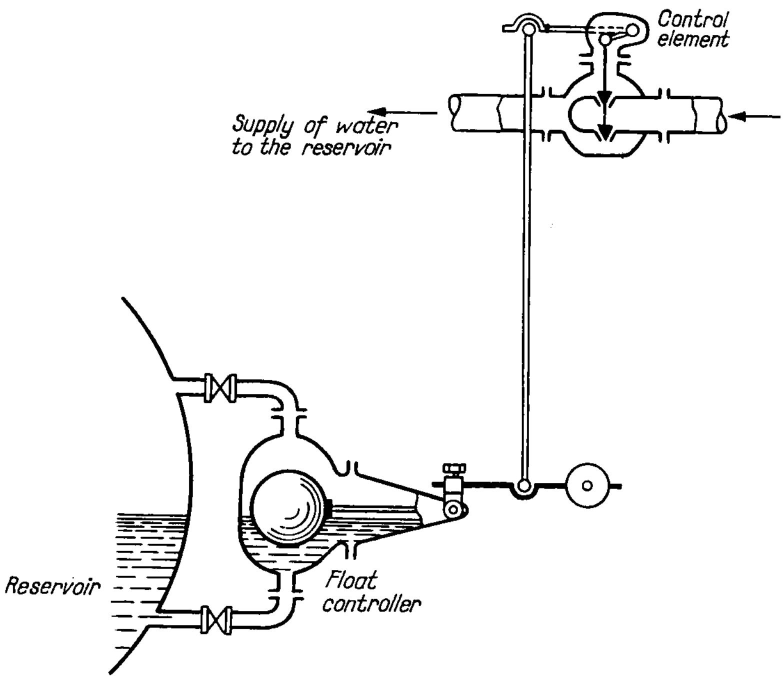


FIG. 6

the carburettors of automobile engines (Fig. 5) to large-scale controllers in large industrial reservoirs (Fig. 6).

The range of application of direct action pressure controllers is just as wide. Any reducing valve or gas reducer (Fig. 7) gives, on a small scale, an example of such a controller. And, as examples of large controllers which sometimes develop forces measuring hundreds of kilograms, and even tons, we may take the mains automatic controlling valves installed in large-scale gas mains (Fig. 8).

The controllers we have described are all static. If a static controller is disconnected from the control element and the value of the

controlled parameter transmitted to it is changed, then the equilibrium position of the regulating unit which is not connected to the control element will also change. Thus, for example, if we disconnect

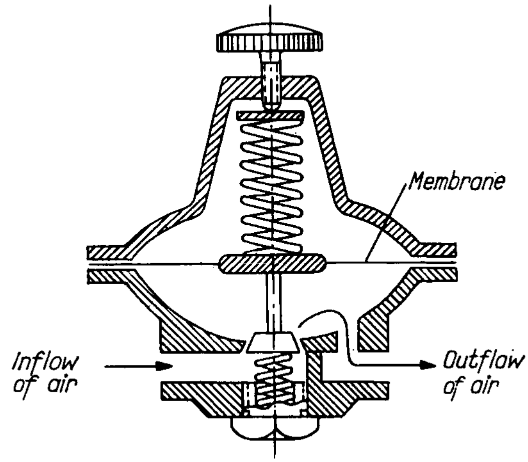


FIG. 7

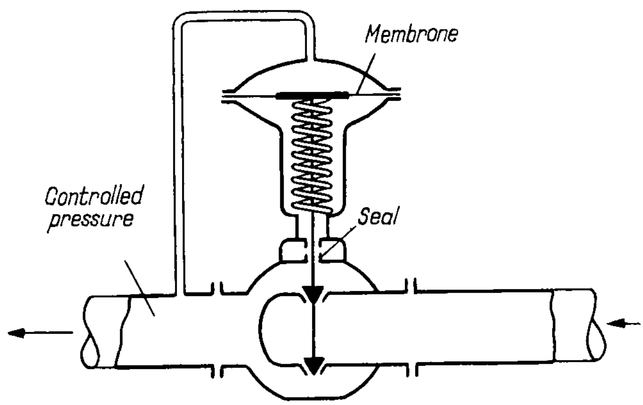


FIG. 8

the coupler of the controller shown in Figs. 2 or 3 from the control element, then, when the angular velocity changes to each new value, there will correspond a completely defined position of the coupler.

When such controllers are used a static error is unavoidable: to various loads on the controlled object there correspond various values of the controlled parameter.

A typical example of an astatic controller is a piston pressure controller, where the piston is loaded by a weight and not by a spring (Fig. 9).

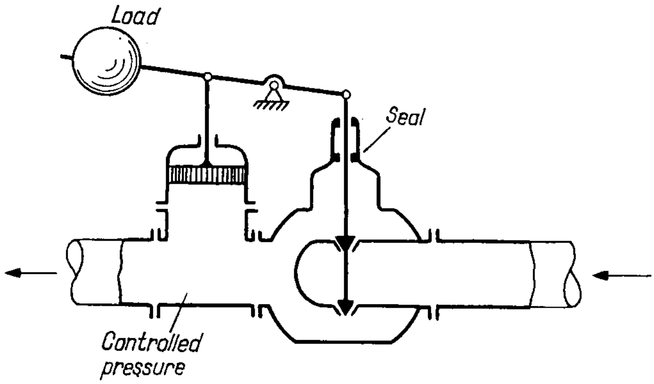


FIG. 9

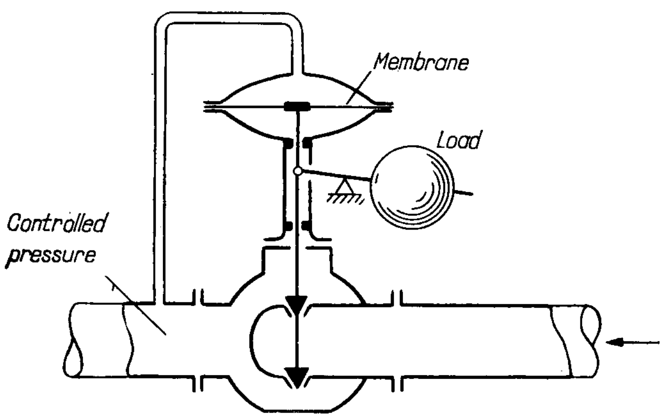


FIG. 10

Equilibrium is possible only for a completely defined value of the controlled pressure and consequently only this value of the pressure can be obtained for any equilibrium position of the control element.

Astatic control also occurs in membrane pressure controllers with a heavy load (Fig. 10), since the force transmitted from the weight to the rod, when within the limits of the operating stroke of the rod,

is almost independent of its position. Centrifugal and level controllers may also be made astatic.

Astatic controllers usually require some means of stabilization. Derivative action is most frequently used for this purpose. Figure 4 gave an example of the construction of a centrifugal direct action controller with derivative action. Figure 11 gives an example of a pressure controller of this type. The chambers A and B are connected by a calibrated opening in the piston. The differential pressure on

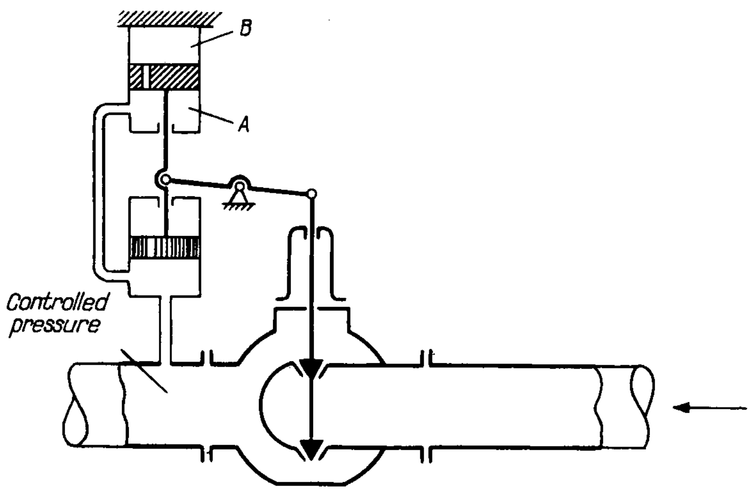


FIG. 11

the piston depends on the rate of change of the pressure in A: it is almost non-existent for a slow pressure change and it is large when the pressure changes quickly. The force produced in both pistons is compounded in the rod and is balanced by the force produced at the valve.

In order to stabilize astatic controllers we sometimes introduce "time staticism", constructing the controller so that, without a damper, it is astatic, and connecting the controller coupler with a spring whose other end is fixed in a hydraulic damper with a very small calibrated opening* (Fig. 12). When equilibrium has been established, the damper piston moves under the action of the spring until the force in the spring

* A centrifugal controller is shown in Fig. 12. We assume that the alignment of the lever arms is chosen so that the characteristic of the controller is almost astatic.

is removed and the static error is eliminated. This unit is the simplest example of a *floating control mechanism,* about which we shall say more later.

We now consider an example of an electric direct action controller.

Figure 13 gives the scheme of the static control of the potential of a generator by means of changing the current in its excitation coil.

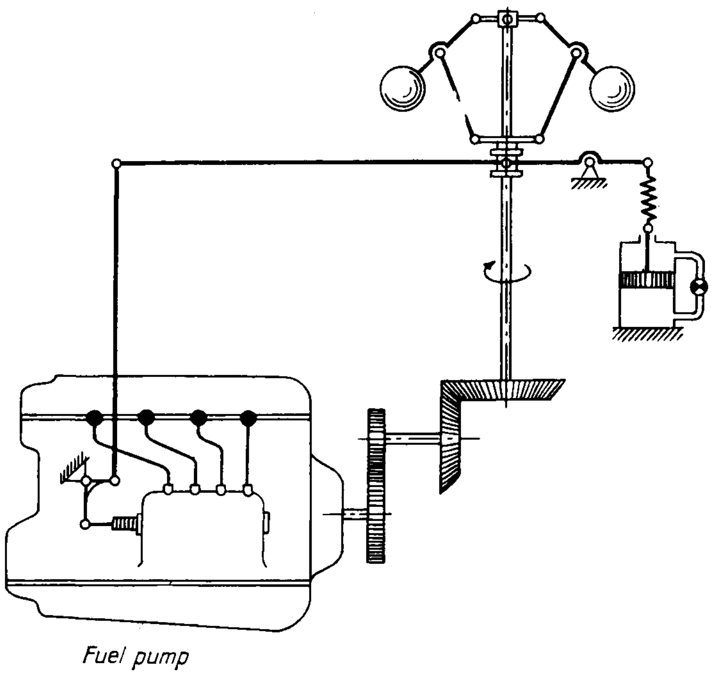


FIG. 12

To maintain the desired potential, the necessary change in the excitation current is made by changing the resistance of the carbon piles, which depend on the tension in the spring and on the force produced by an electromagnet to whose coil the difference between the potential of the generator and the standard potential U_0 is applied. When the generator potential changes, so does the current in the excitation coil of the electromagnet, the pressure applied to the carbon piles and the excitation current of the generator, so that the necessary decrease in the deviation of the potential from the desired value is obtained.

This controller is a direct action controller, since the energy which is required to produce the controlling action (the pressure

applied to the carbon piles) is here obtained from the controlled object via the sensor.

The basic advantage of direct action controllers is their exceptionally simple construction. Experience in their use has shown, however, that they also possess many defects.

In direct action controllers it is unavoidably necessary to increase the size of the moving masses in order to ensure the large translatory

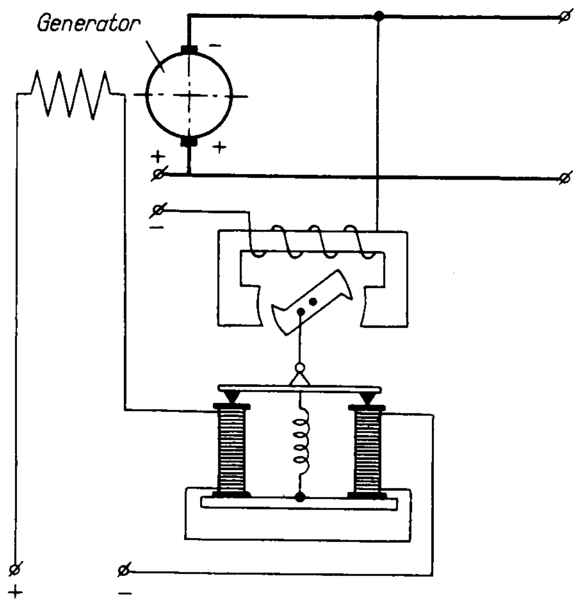


FIG. 13

forces required. As a result, the stability of the controllers is decreased. Large forces increase friction, and so the insensitivity increases.

In some cases these forces reach hundreds of kilograms, and sometimes several tons, or even tens of tons. Moreover, difficulties arise in the choice of measurers for many parameters (such as temperature) which could produce any large translatory force with a sufficiently high accuracy.

This leads naturally to a simple idea: to place an amplifier between the measurer and the control element so that the necessary translating force can be produced in the amplifier by transferring energy from outside, and a low power measurer may then control the amplifier.

3. Astatic Controllers of Indirect Action

A typical example of an astatic indirect action controller with a hydraulic amplifier is shown in Fig. 14. In equilibrium the valve returns to the same (neutral) position independently of the position of the piston and of the control element moved by it. The "pointer" of the

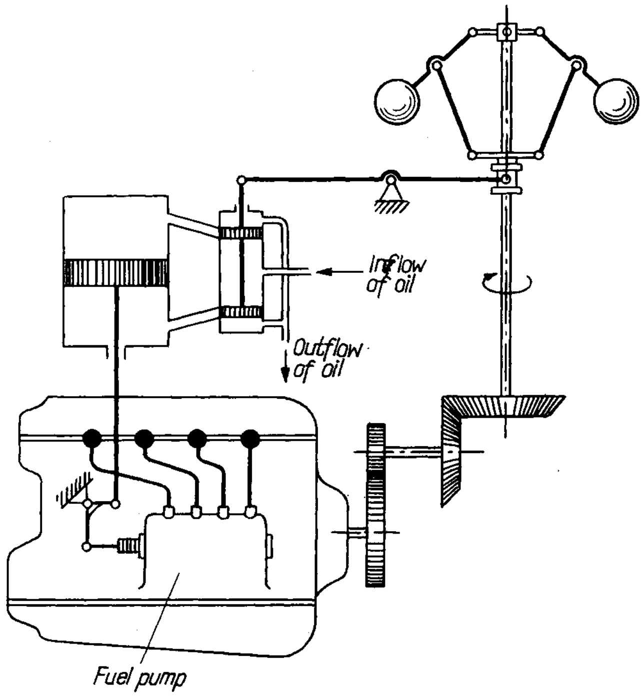


Fig. 14

sensor* returns accordingly to the same position, and hence for any load the value of the controlled parameter which is set up will be the same.

Figure 15 shows a similar controller, with jet distribution. The fact that in the equilibrium position the jet pipe must be set exactly mid-way between the receiving nozzles ensures that the controller is astatic.

* By the pointer of the sensor is meant, conventionally, some part of it whose position is proportional to the position of the point of the sensor by which it is connected to the preceding element in the control system. In sensors with indicators or recording systems, this could literally be a pointer, or in "blind" devices, any point of their moving system.

An example of a controller with a pneumatic astatic amplifier is shown in Fig. 16. The air pressure in chamber A between the fixed restriction and the nozzle covered over by the flapper depends on the position of the flapper. For normal sized nozzles the working stroke of the flapper does not exceed 0·1 mm. Thus, a very slight displacement

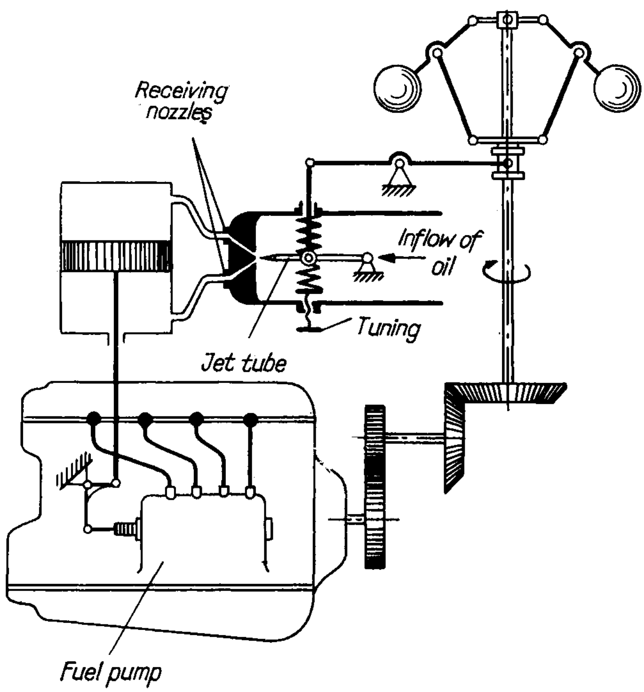


Fig. 15

of the flapper is sufficient to make the air pressure in A fall to almost atmospheric pressure or rise to the pressure in the supply line. When equilibrium has been reached in the system, the flapper returns to roughly the same position (within the limits of the very small working stroke). Hence in any equilibrium position, the value of the controlled parameter will be the same.

The pressure change in the chamber A controls the output from a second amplifier, namely the pressure in chamber B. The air is conducted under this pressure to the membrane control unit.

The speed control system for an electrical machine shown in Fig. 17 represents an astatic controller of indirect action like those

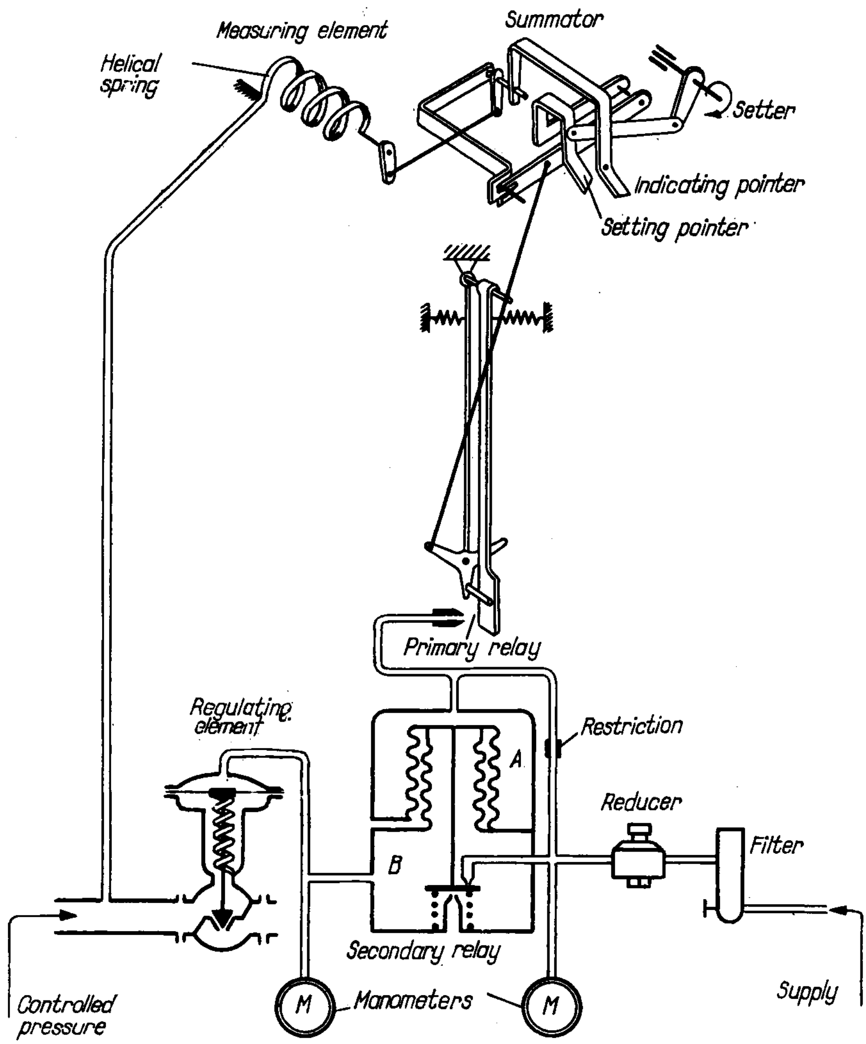


Fig. 16

considered above. The signal characterizing the value of the controlled quantity, the speed of the machine, is obtained here with the help of a tachogenerator. The potential of the tachogenerator is compared with the standard potential U_0, which characterizes the desired machine velocity. The potential difference is communicated to the input of an amplidyne whose output is connected to a control motor,

which alters the resistance in the excitation coil circuit of the electric motor. Equilibrium in the system may, clearly, only take place when there is no potential at the input of the amplifier, and hence no deviation of the machine velocity from the desired value.

Figure 18 shows the circuit of an automatic temperature control system using an astatic indirect action controller. The temperature

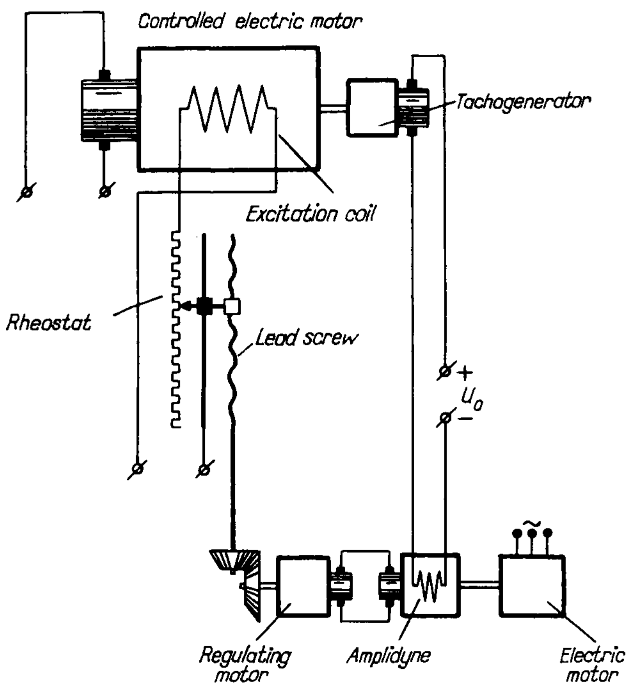


FIG. 17

of the controlled object, the heat exchanger, is measured by a resistance thermometer in the electrical bridge circuit which is balanced by a resistance R_0 when the temperature is at the desired value. When the controlled temperature deviates from this value, a potential U will be produced in the measuring diagonals of the bridge, its sign and magnitude determined by the sign and magnitude of the temperature deviation. This potential, amplified by an electronic amplifier, is transmitted to the input (control coil) of a magnetic amplifier, which is in one of the arms of the inductive bridge and is transformed into a variable current potential. The bridge circuit supplies the control

coil of a two-phase asynchronous motor with a short-circuited rotor moving the control tap which increases or reduces the supply of heat.*

In static direct action controllers there is a one-to-one correspondence between each position of the sensor pointer and the position of the control element. In contrast, in astatic controllers, the position of the sensor pointer determines the displacement velocity of the

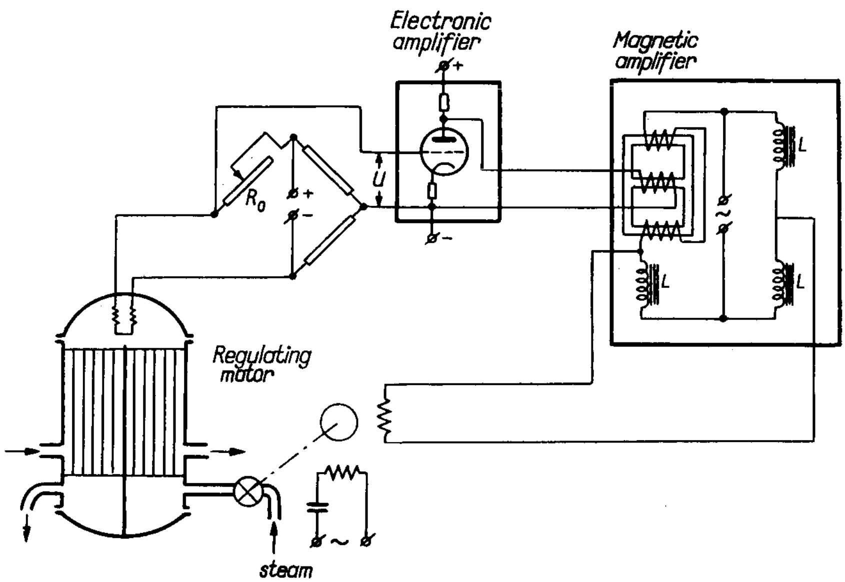


FIG. 18

control element but not its position. The examples given above included both those controllers in which the velocity of the control element increases smoothly as the deviation of the sensor element increases (Fig. 17) and relay controllers, where even a very small deviation of the pointer causes the control element to move with a constant velocity (Fig. 16). Depending on its properties any astatic servomotor can approximate to a relay servomotor if the working stroke of the sensor is decreased.

The introduction of an amplifier enables us to use high quality measuring instruments with a small moving mass which develop

* For the sake of clarity, the amplifier circuits shown here are simplified, being single-stage, and with only one controlled arm in the bridge. In actual circuits two- and three-stage electronic amplifiers are used, and two, or sometimes all four, arms of the bridge are controlled.

small forces.* But because of this introduction of an astatic amplifier, the one-to-one correspondence between the position of the pointer and that of the control element is destroyed, and this encourages instability.

The invention of feedback enabled us both to use an amplifier and to retain the one-valued correspondence between the position

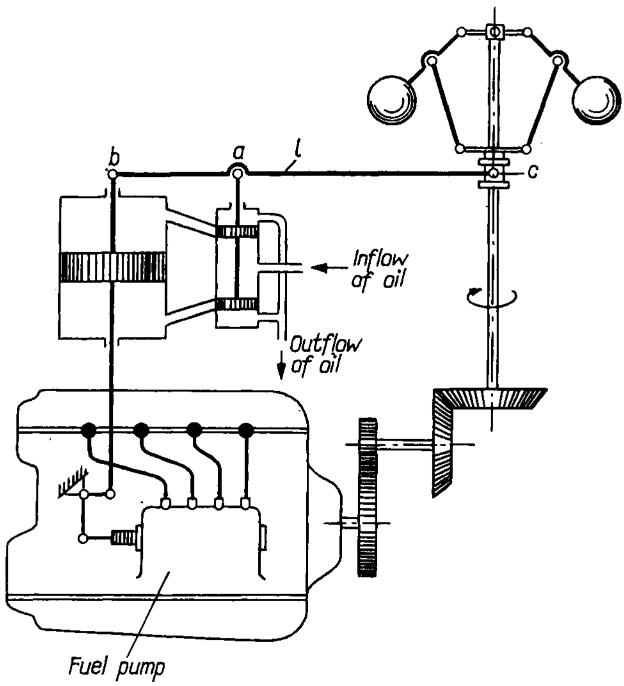


FIG. 19

of the control element and that of the pointer of the sensor. The development of contemporary control techniques is basically connected with this invention.

4. Rigid (Proportional) Feedback. Static Indirect Action Controllers

The controllers which we examined in the previous section are shown in Figs. 19—23 with the addition of rigid feedback. Because of the feedback, the position of the controlling unit of the amplifier

* Up to a few grams, but sometimes only a fraction of a gram.

depends not only on the controller sensor but also on the deflection in the regulating unit.

In the simplest case (Fig. 19) the feedback is realized by placing the support of the lever l on the piston rod. Although the valve (and

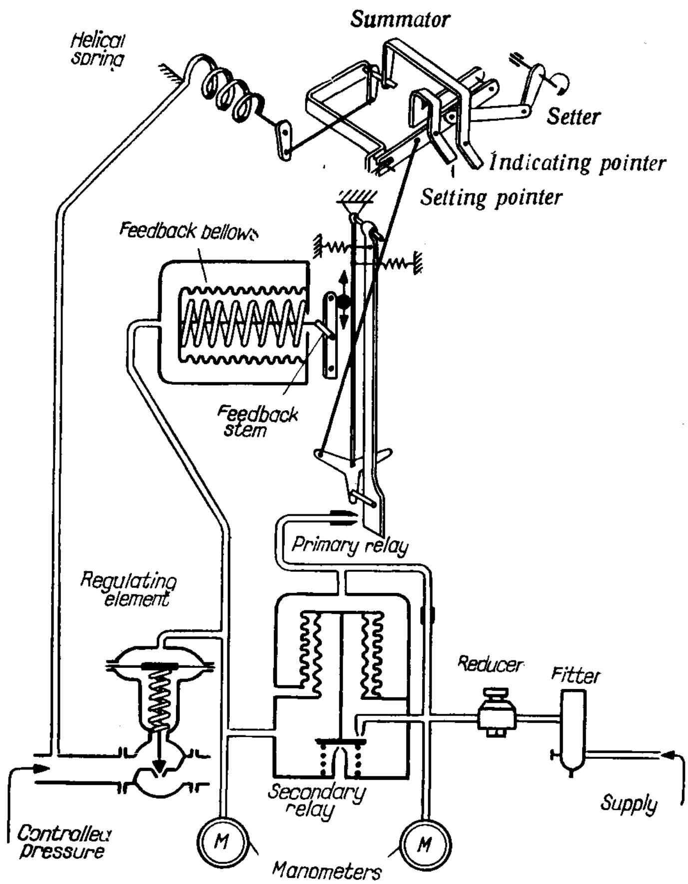


FIG. 20

the hinge a) take up a neutral position in equilibrium, as before, the position of the piston determines the displacement of the hinge b and, therefore, of the hinge c which is connected to the sensor. Due to this, the deviation of the sensor pointer is proportional to the displacement of the piston and, of course, the control element also. The feedback restores the simple correspondence between the deflection

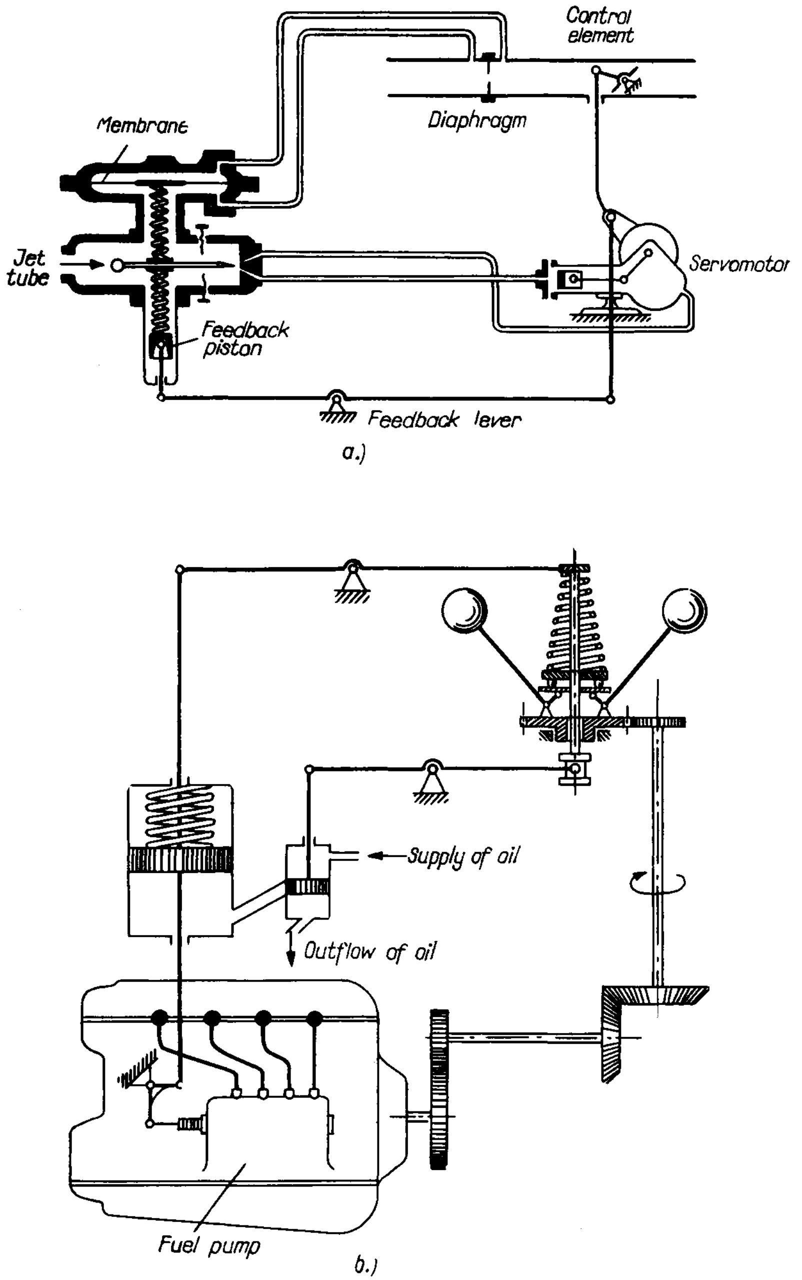

Control element
Diaphragm
Membrane
Jet tube
Servomotor
Feedback piston
Feedback lever
a.)
Supply of oil
Outflow of oil
Fuel pump
b.)

FIG. 21

of the sensor pointer and the position of the control element. This furthers the stability of the system but makes it static, that is, once again a static error is introduced: to each value of the load on the object there corresponds its own value of the controlled parameter.

In the example just considered, the feedback was realized by a rigid lever. Feedback in the pneumatic controller shown in Fig. 20

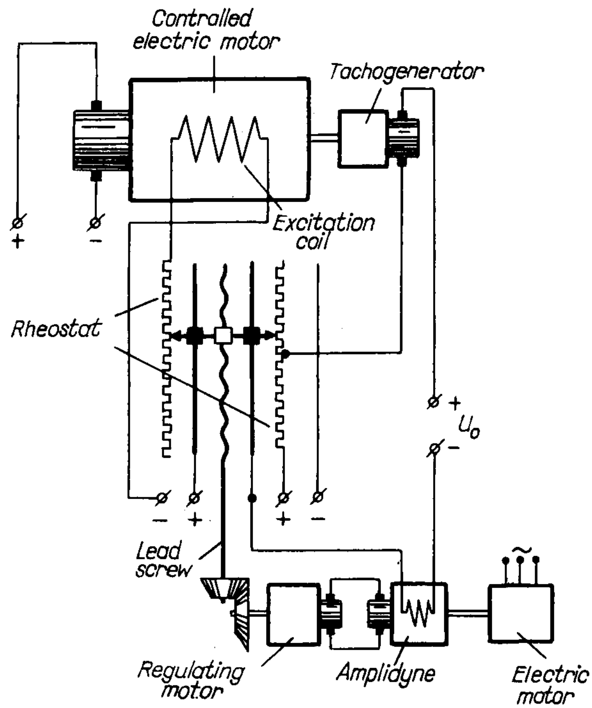


FIG. 22

acts in exactly the same way. In this case the position of the flapper depends not only on the deviation of the sensor pointer but also on the tension in the feedback bellows to which the air directed to the regulating unit is brought. In the position of equilibrium the tension in the bellows is proportional to the position of the control element.

The feedback in a jet controller (Fig. 21*a*) is constructed somewhat differently. Here the jet tube is squeezed between two springs. One of them is compressed by the sensor of the controller, and the other by the feedback piston. The tube acts as if it were scales

weighing the difference between the tension of the springs. In the position of equilibrium the jet tube returns to the neutral position, but the tension in the spring above the tube must then be equal to the tension in the spring below the tube, i. e. the value of the controlled parameter in the position of equilibrium is determined by the position of the control element.

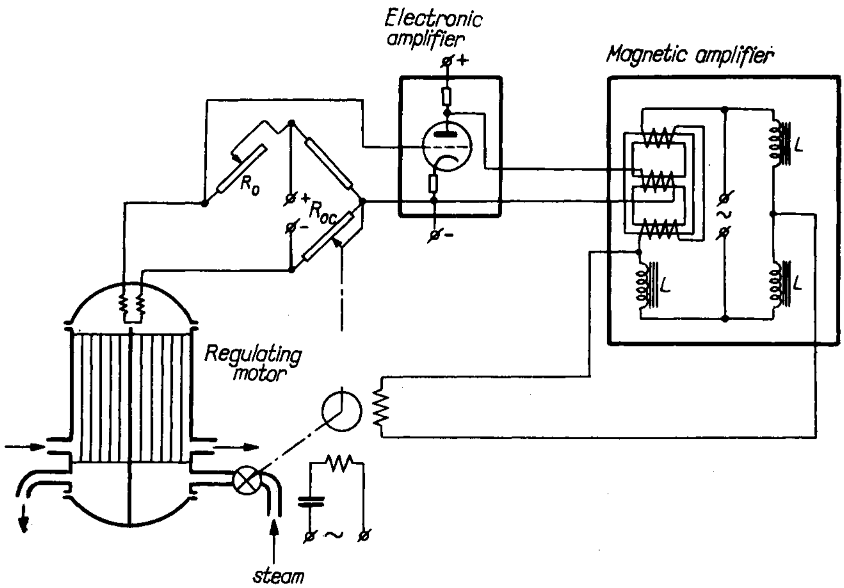


FIG. 23

One more example of this kind, for a valve distribution, is shown in Fig. 21*b*. In this controller the feedback compresses the spring which directly acts on the sensor.

The controller of the speed of an electric motor, shown in Fig. 22, is an example of an electric indirect action controller with rigid feedback. Here the feedback signal is obtained from a potentiometer whose slide is rigidly connected to the cursor of the rheostat in the excitation coil circuit of the motor. The input of the amplidyne receives, in addition to the error signal, a feedback signal. Of course, here too the improvement achieved in the dynamic properties of the control system is bought at the cost of introducing a residual deviation in the controlled quantity, in so far as in this system an equilibrium

state can only be attained when there is a deviation in the motor speed from the desired value such that the difference in potential of the tachogenerator and the standard potential U_0 balances the potential of the feedback signal.

Figure 23 shows the simplified circuit of a temperature controller with rigid feedback. Here, in contrast to Fig. 18, the displacement of the control element causes a change in the resistance of one of the arms (R_{oc}) of the measuring bridge, as a result of which there is a change in the temperature for which the bridge balances. Thus there

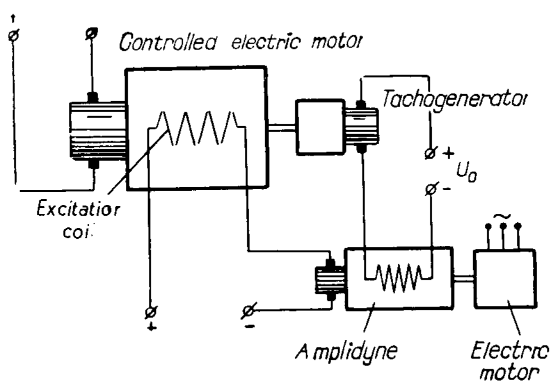


FIG. 24

corresponds one equilibrium value of the controlled temperature to each value of the load and therefore to each position of the control element.

In the system used to control the speed of an electric motor, shown in Fig. 24, there is no special feedback element, although this system is also static, since all the elements in the control circuit are static.

Figure 25 shows the circuit of an automatic voltage regulator, which makes use of the influence of the generator excitation on the generated potential. The desired value of the generator potential is here attained by changing the position of the crossarm of the plane controller and at the same time changing the resistance in the excitation coil circuit of the generator and the current flowing through a given coil B_1 of the amplidyne. Through the coil B_2 of the amplidyne there flows a current proportional to the generator potential. The

difference between the ampere turns in B_1 and B_2 determines the value of the potential at the amplifier output, and therefore also the potentials of the exciter and the generator.

All the elements of this controller are static, and the controller as a whole is static. This is obvious, and immediate, since a change in the steady value of the control action is here necessarily connected with the change in the deviation of the potential from the desired

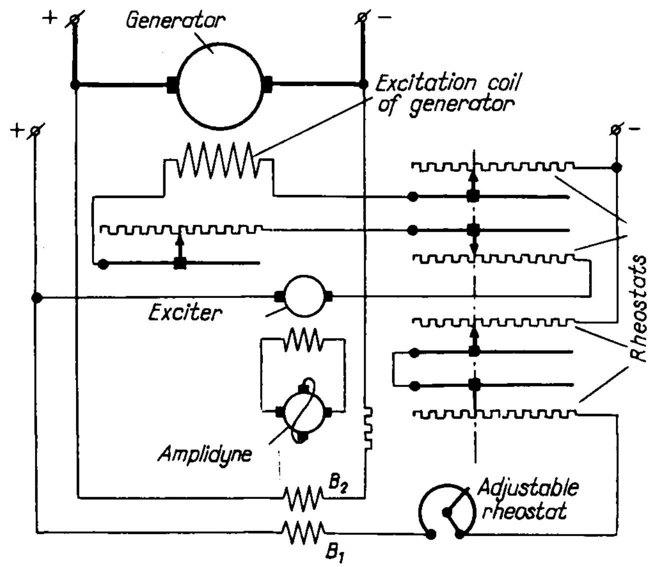


FIG. 25

value, and to each value of the load current there corresponds a different value of the potential.

It was shown above that the feedback which restores the one-valued correspondence between the positions of the sensor pointer and of the control element promotes the stability of the system but also introduces a static error. The idea of so constructing the feedback that it can act only during the control process and is removed when the system approaches the equilibrium state naturally arose. Feedback constructed in this way enables us to combine the advantages of static and of astatic controllers. During the control process the controller operates as if it were static, but, gradually, according to

how much of the feedback effect is removed, the value of the parameter maintained by the controller returns to the same value.

Feedback of this kind is given the name "floating".

5. Floating Feedback

Figure 26 shows a hydraulic valve controller with floating feedback. When the piston of the hydraulic amplifier moves, the point *a* of the feedback lever also moves, since the oil filling the hydraulic damper does not have time to flow through its small calibrated opening. But when the amplifier piston is motionless, the oil in the damper slowly flows from one space to the other under the action of the floating component, and the point *b* always returns, slowly, to its initial position. The point *a* returns to its initial position, since in the equilibrium position the valve must close the port. In the equilibrium position, therefore, the point also returns to the old value of the controlled quantity.

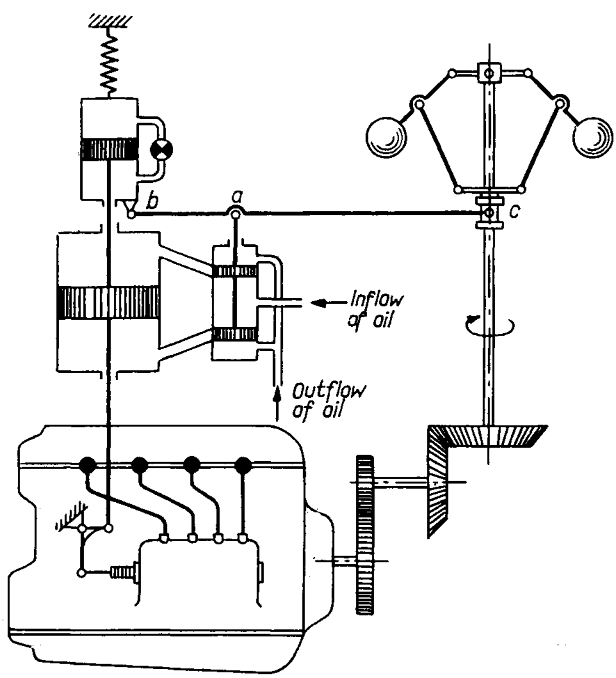


FIG. 26

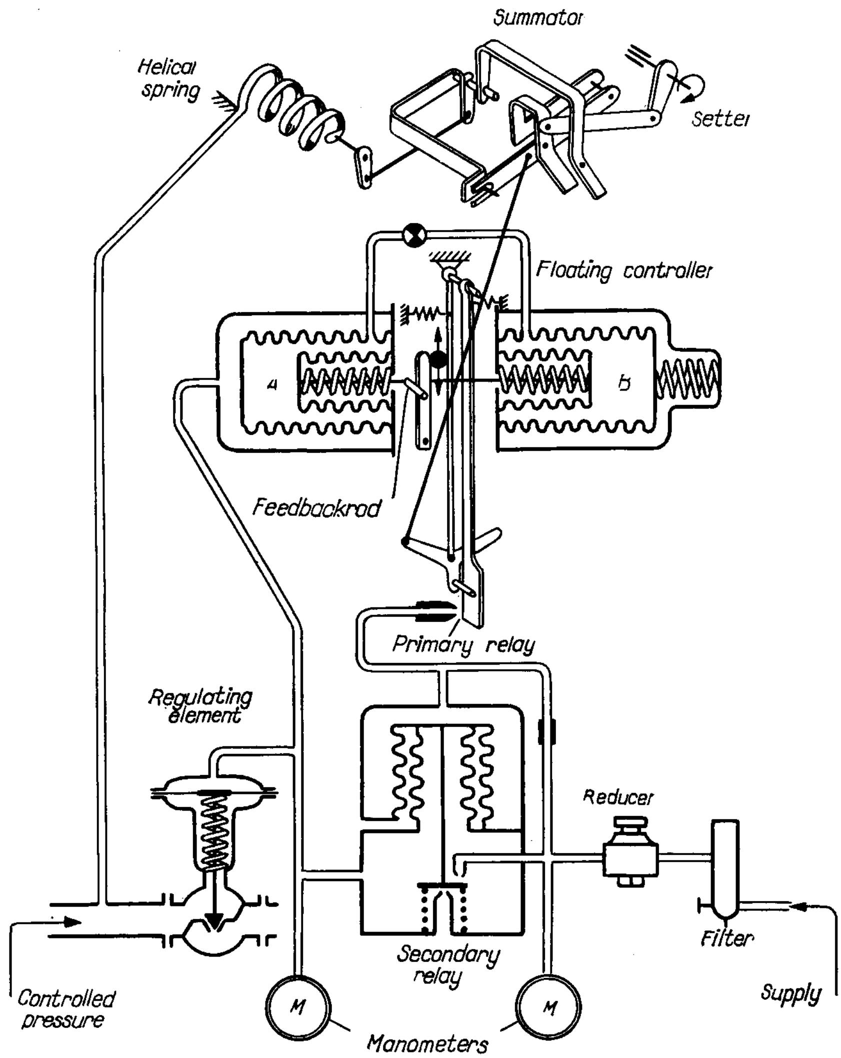


FIG. 27

An example of a pneumatic floating controller is given in Fig. 27. In contrast to the controller of Fig. 20, the feedback effect is here gradually removed as a result of the flow of the liquid (toluene) out of the bellows A into the bellows B through the controlled throttle. When the pressure in A is equal to that in B the feedback rod balances and returns to its initial position.

The floating control unit of an aggregate unifissionary system (AUS) is shown in Fig. 28. A pressure proportional to the instantaneous value of the controlled quantity is led along the channel O, and along the channel O_1 is conducted a pressure proportional to the desired value of the controlled quantity.

When the controlled quantity deviates from the desired value a pressure difference occurs, which displaces the membrane and changes

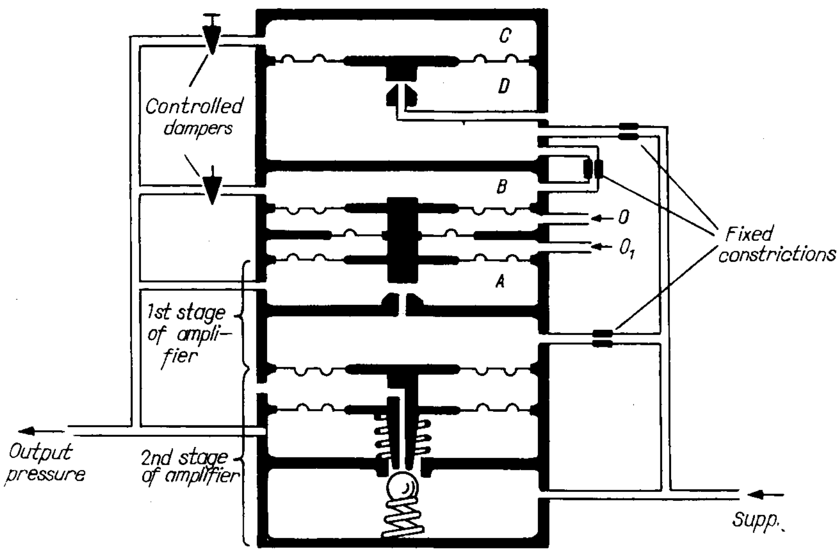


FIG. 28

the position of the flapper fixed to the membrane in the first stage of the amplifier. The pressure set up at the output of the second stage of the pneumatic amplifier is led underneath the membrane to the chamber A; total feedback is thus put into effect. In order to control the effect of the feedback this same pressure is brought to chamber B through a controlled throttle. If the pressure in chambers A and B were equal, then the feedback effect would be removed. But, because of the presence of the throttle, the pressure in B is less than that in A, and by changing the size of the throttle opening, we can change the effectiveness of the feedback. The pressure from A is led through another controlled throttle to the empty chamber C. In time the pressure in C always becomes the same as that in A, but the time taken

for these pressures to become equal can be changed by altering the section of the throttle. The resulting system maintains the pressure in chamber D equal to the pressure in C. Therefore the pressure in all the chambers (A, B, C and D) is gradually equalized and the feedback effect is thereby removed.

The hydraulic and pneumatic controllers with floating feedback which we have just described are also used in electrical controllers.

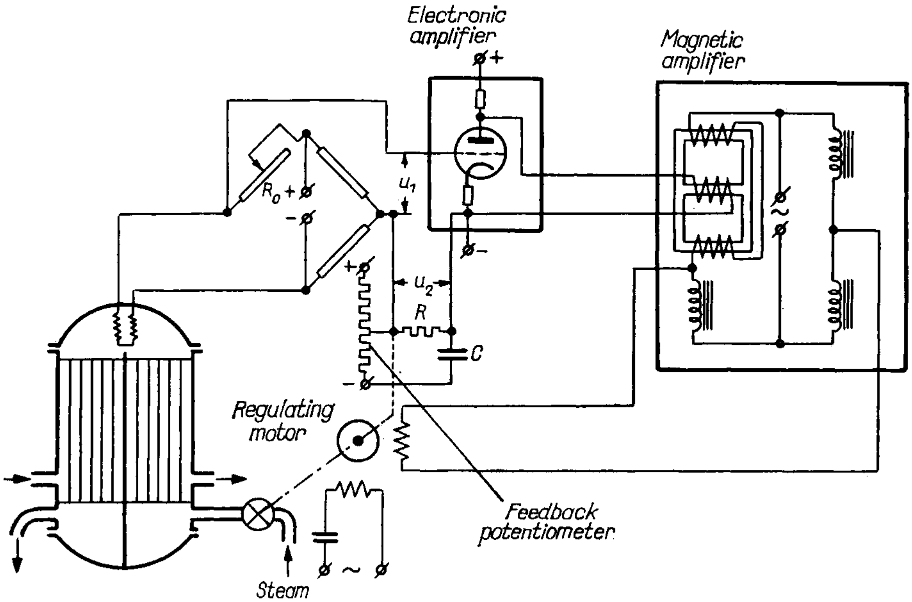


FIG. 29

Thus, for example, a floating system of temperature control can be obtained by substituting the rigid transmission from the control element to the feedback resistance arm in Fig. 23 by a transmission which includes an air or liquid floating component. Floating feedback can, however, also be realized by purely electrical methods, as shown in Fig. 29. In this system the feedback signal is derived from a potentiometer and is supplied to the control system via a network containing a resistance R and a capacitance C. During the transient state of the system this signal will, together with the deviation signal, act on the subsequent elements of the system, although if the control element were motionless, which never happens, the feedback signal would die away at a rate depending on the charge on the condenser of C, and

would always be zero in the steady state. It is obvious here that the feedback will not cause residual deflexions in the controlled quantity to appear.

The potential controller shown in Fig. 30 is an example of a floating controller constructed by introducing an astatic element into a static controller. Here, when the potential becomes different from the desired value, the coil B_1 of the amplidyne immediately becomes

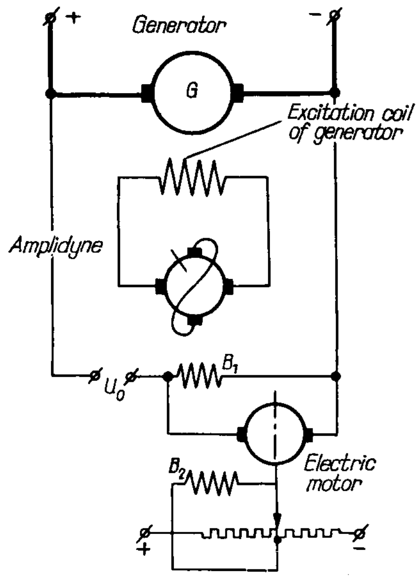


Fig. 30

effective. Simultaneously the potential of the coil B_2 begins to change at a speed proportional to the deviation in the potential. Equilibrium in the system occurs when the potential deviation becomes equal to zero. The potential of B_1 then also returns to zero, and that of B_2 becomes equal to the potential necessary to maintain the generator potential at its desired level, and the motor which displaces the potentiometer slider stops.

The amplidyne potential controller shown in Fig. 31 is another example of a controller with flexible (floating) feedback. In contrast to the static potential controller examined above, here the flexible feedback introduced into the control system is realized with the help of a stabilizing transformer. The current in coil B_3 of the amplidyne is

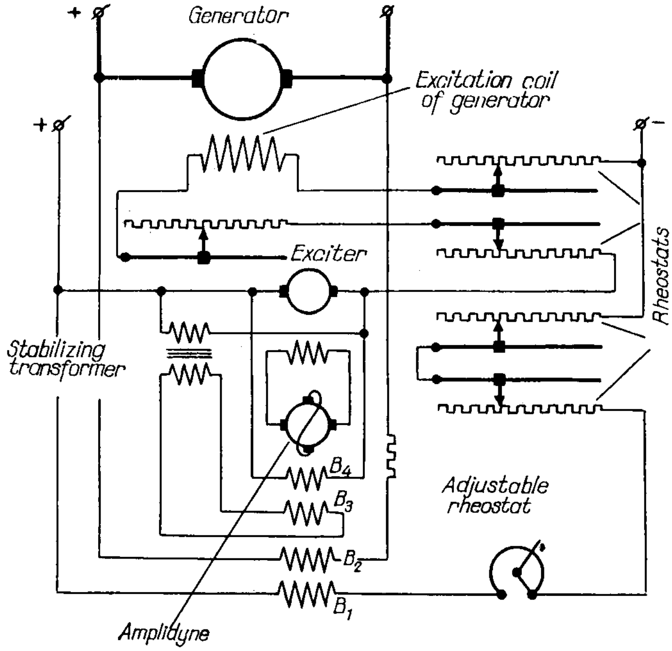


FIG. 31

proportional to the rate of change of the exciter potential, and the primary coil of the transformer is connected to the exciter. In the steady state the exciter potential is invariable and the current in B_3, supplied by the stablizing transformer, is equal to zero. Thus, this

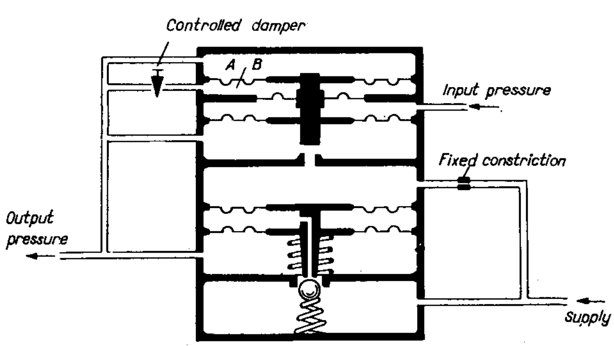


FIG. 32

feedback, like all floating feedbacks, does not increase the steady deviations of the controlled quantity from the desired value.

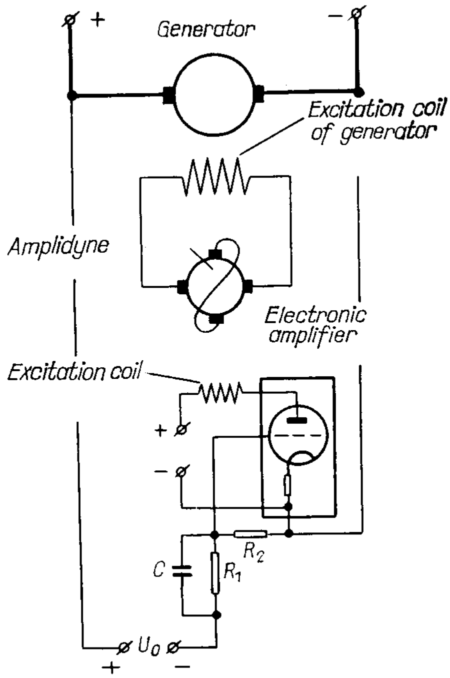


FIG. 33

6. Derivative Actions

In Section 2 of this chapter we gave the most simple example of a direct action controller using derivative action. In indirect action controllers derivative action is widely used as a method of stabilization, the unit producing the action being placed at various points of the circuit.

In addition to mechanisms which directly measure the derivative from the controlled parameter* differentiating units are widely used.

An example of a pneumatic unit producing derivative action is shown in Fig. 32, which gives the scheme of the anticipation unit of an aggregate unifissionary system. For a slow change in the input

* An example of such a mechanism is a tachometer when the controlled parameter is the angle of revolution of any shaft.

pressure the output pressure differs little from it, but if the input pressure changes rapidly then output pressure will be larger than it and, moreover, the higher the velocity of change of the input pressure the larger will the output pressure be. Thus, the output pressure depends both on the magnitude of the input pressure and on the speed at which this pressure is changing. The amplifying part of this unit is the same as that in the controller of Fig. 28.

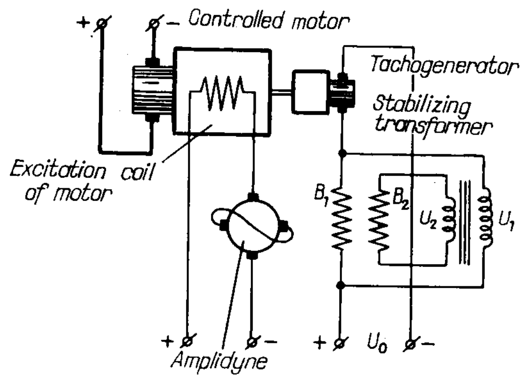


FIG. 34

In electrical controllers derivative action is most often introduced into the control system by means of a so called differentiating or *RC* circuit, the arrangement of which is shown in Fig. 33, where we show in its simplest form a system of automatic voltage control using derivative action. The signal of potential deviation is sent to the input of the differentiating circuit which consists of a capacitance C and resistors R_1 and R_2. A signal proportional to the potential deviation and to the time derivative of this deviation is then obtained across R_2.

The current through the resistor R_2 may be assumed to be approximately equal to the sum of the independent currents I_R, flowing through the resistor R_1, and I_C, flowing through C. The current I_R is obviously proportional to the potential at the input and I_C is proportional to the rate of change of this potential. The potential across R_2, which is is used as the output signal, being proportional to the total current through R_2, therefore contains components proportional to the input signal and to its rate of change.

Figure 34 gives the circuit of a controller of the speed of an electrical machine, where the derivative action signal is obtained by using a stabilizing transformer. The potential U_1, proportional to the deviation of the speed of the machine is applied to the primary coil of the transformer. The potential of the secondary coil, U_2, which is proportional to the rate of change of U_1 is applied to the coil B_2 of the amplidyne, as a result of which the control action becomes proportional not only to the deviation of the speed, but also to the derivative of this deviation.

7. Multiple Control

An example of a system of multiple control is given in Fig. 35. This is a hydraulic controller with connexions joined after first-stage amplification. In this circuit each sensor controls only one amplifier.

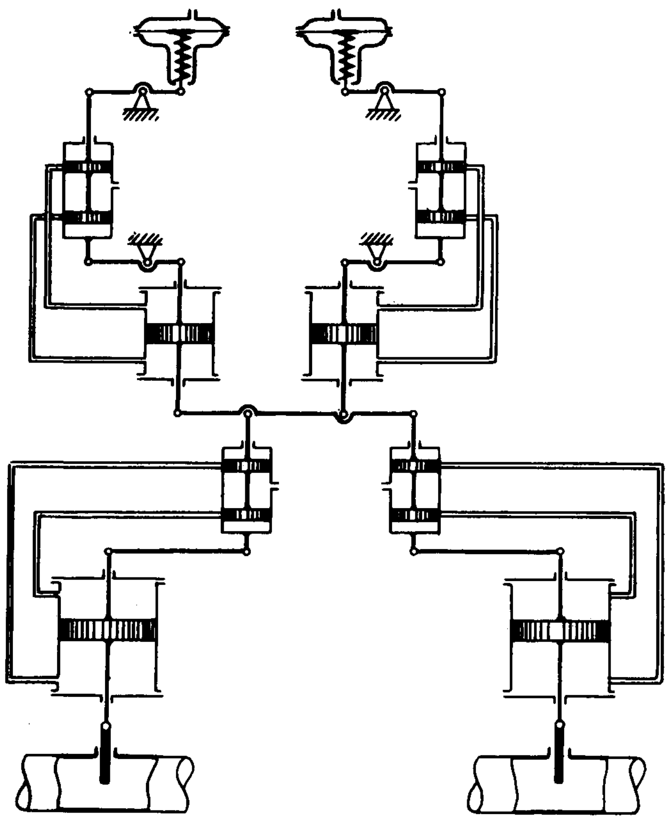


FIG. 35

Each of the second-stage amplifiers is controlled by the two first-stage amplifiers.

Figure 36 shows a similar circuit with the connexions joined directly after the sensors.

Here each sensor controls all the hydraulic amplifiers, but any one amplifier is connected to only one control element.

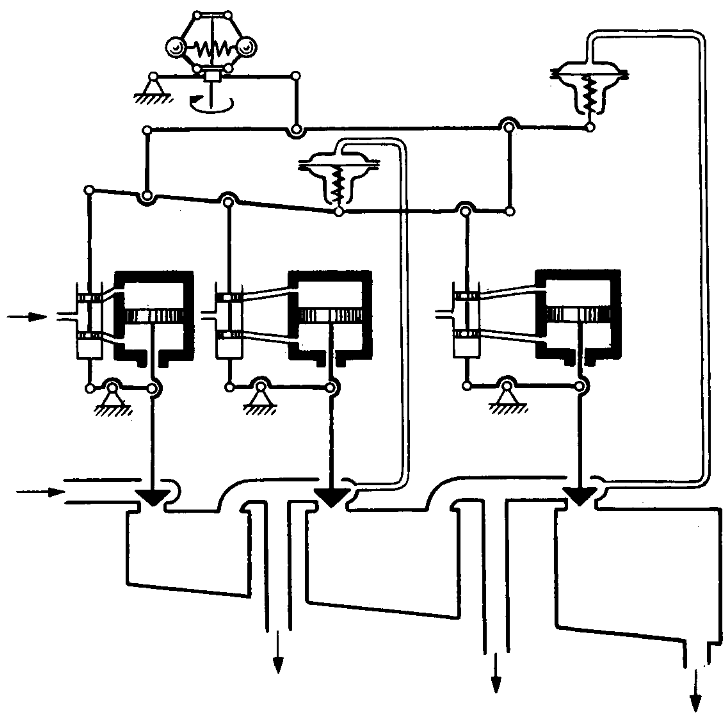


FIG. 36

Another example of multiple control is afforded by a system of control with two regulating units leading to two control elements (Fig. 37).

The feedback potentiometers shown in the circuit are connected in such a way that a displacement in one of the control elements causes a corresponding displacement in the other.

Another example of multiple control is shown in Fig. 38.

In this circuit the position of the setter of the steam flow controller depends on the sensor signal reacting to the temperature change.

8. Two-position (Oscillatory) Controllers

The controllers described above all possessed a position of equilibrium. This means that for each value of the load on the controlled object and for each position of the setting mechanism of the controller there existed a position for all elements in the controller in which they and the controlled object were in equilibrium.* But we can also construct controllers so that the system as a whole which contains them does not, in principle, possess an equilibrium position. Such controllers

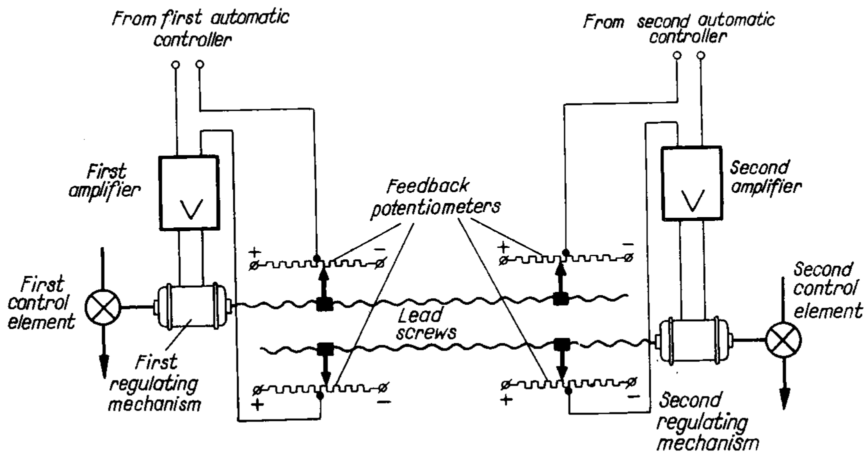


FIG. 37

set the control element in only one of two fixed positions (for example, "completely closed" or "completely open") or else move it with one of two fixed velocities (for example, "full speed forward" or "full speed backward"). To obtain equilibrium of the controlled object and the controller it would be necessary to set the control element in an intermediate position, determined by the load, but these controllers cannot do this and they move the control element when there is a change in the sign of the difference between the actual value of the controlled parameter and the desired value set by the setting mechanism. As a result the operating state consists of undamped oscillations: the controlled parameter oscillates about the desired value,

* This equilibrium can be stable or unstable (this depends on the tuning of the controller) but it exists in principle.

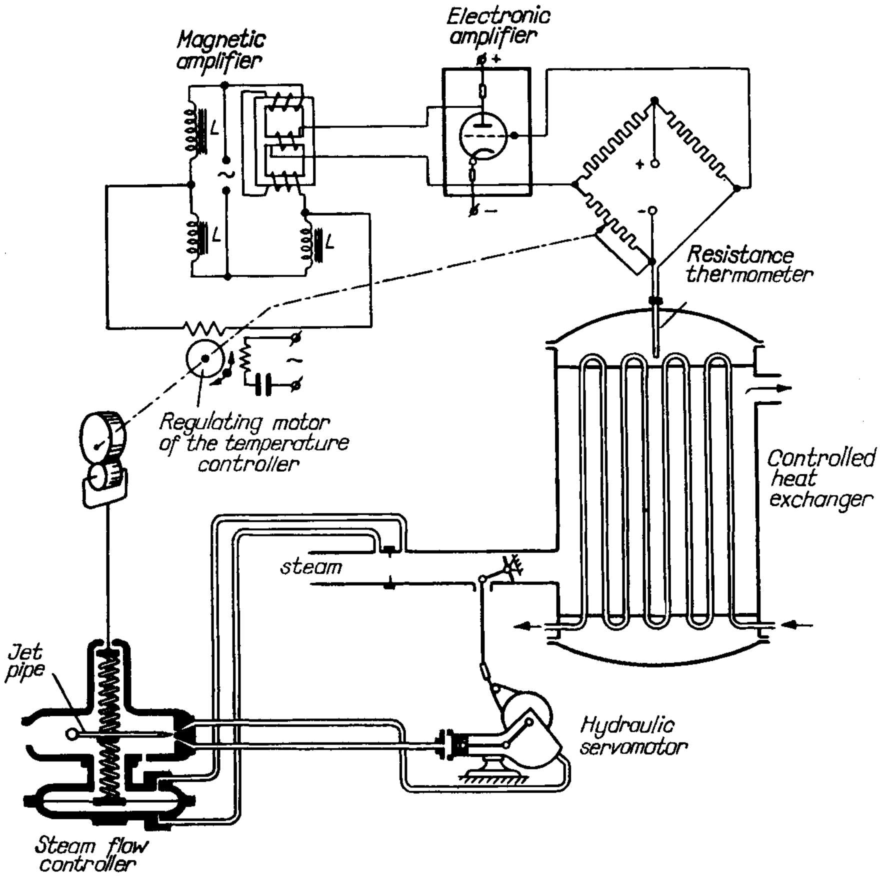


FIG. 38

and the control element or its velocity continuously "jumps" from one extreme position to the other.

Controllers of this kind are called *two-position* or *oscillatory*. The simplest example of a two-position temperature controller is shown in Fig. 39.

The sensor in this controller is a contact mercury thermometer. The position of the contacts sets up the required temperature value. When the mercury column closes the contacts, the heater is completely turned off and when it opens them, the heater is completely switched on. The required temperature value can be set up neither when the heater is completely switched on nor when it is completely

switched off, and requires an intermediate degree of heating. As a result oscillations are set up: the maintained temperature oscillates about the desired value, the mercury column oscillates about the

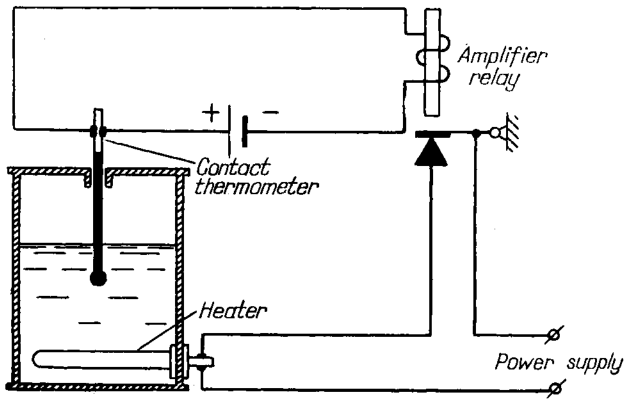


FIG. 39

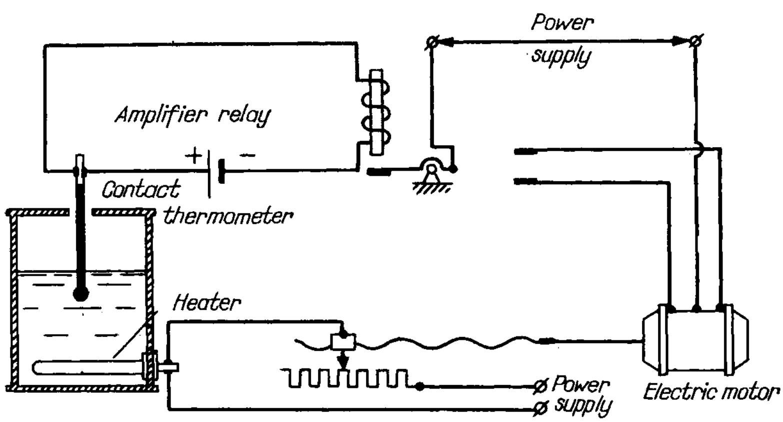


FIG. 40

contact plates and the heater uninterruptedly switches alternately on and off; when the maintained temperature changes, the relation between the time during which the heater is turned on or off (in one oscillatory cycle) changes.

Figure 40 shows a temperature controller with the same sensor. In contrast to the controller considered before (Fig. 39), in this one

when the mercury closes the contact plates, the electric motor which drives a rheostat cursor through a reduction gear with constant velocity is switched on and changes the degree of heat. When the plates open, the direction of rotation of the motor changes.

The circuit of a voltage regulator for an automobile dynamo is shown in Fig. 41.

When the required potential is attained the relay contacts are opened and the additional resistance R_a of the excitation coil circuit

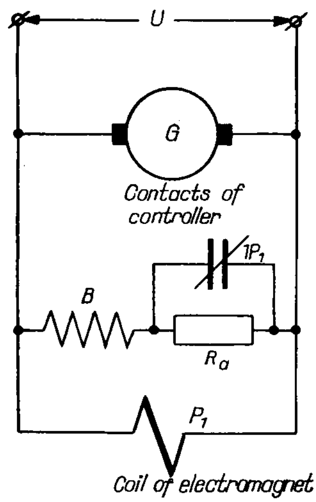


FIG. 41

is included, so that the potential U falls, the resistance R_a is then cut out and so on.

Sometimes a two-position (oscillatory) unit is used as an element of a more complex controller, for example in one of the stages of an amplifier. Besides this unit the control circuit contains continuous units. Their parameters are chosen so that they practically do not react to the oscillations of the oscillating element, but average them and react only to the disturbances of the mean value of the oscillation. Controllers of this kind are sometimes called *combined*.

An example of a combined controller intended for the control of the potential of an electric generator is given in Fig. 42.

If, in this controller, the contact K_2 is prevented from moving, then changes in the potential U cannot influence the operation of the

system, and the controller may then be considered as an ordinary oscillatory controller of the potential at the terminals of the exciter B: it periodically switches the additional resistance R_a in and out of the excitation circuit of the exciter. But in reality the contact K_2 is not motionless, and moves when the potential U_c at the terminals of the main generator changes. Thus, the average value of the oscillation in

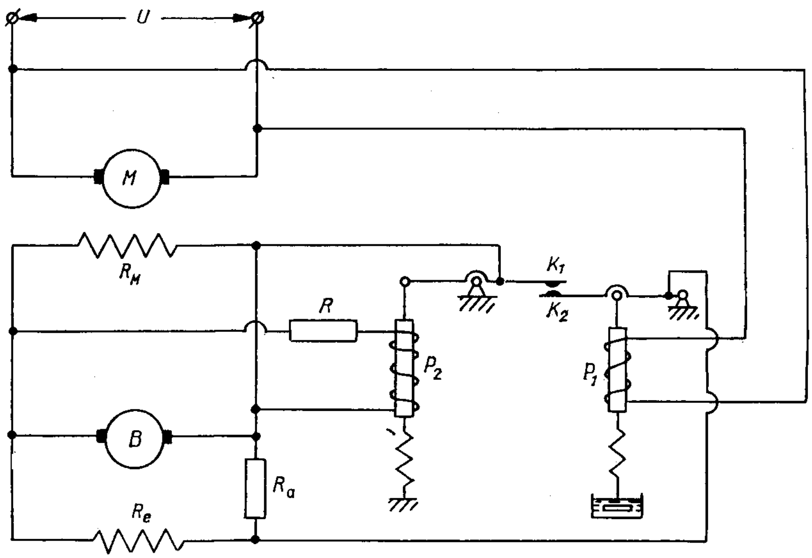


FIG. 42

the oscillatory controller is displaced, since the relative time for which the contact is switched on in each oscillatory cycle is changed.

The system as a whole operates in such a way that the potential U cannot in practice perform oscillations (they are filtered out by the large inductance of the excitation circuit of the generator M), but the oscillatory circuit is essential to the system and plays the part of a special oscillatory regulating unit.

9. Controllers of Discontinuous Action

All the controllers described in the previous sections were *continuous*. Every element acted continuously on the following element. In some cases (when there existed a relay, saturation limits and so on)

this action of one element on the next could, during various stages of the process, remain unchanged but it constantly influenced the conduct of the next element (for example, by determining the moments for closing and opening contacts). In addition to these controllers, there are also controllers of *discontinuous action*, in which the action circuit closes or opens in a manner which is independent of the way the process behaves.

A very simple example of a controller of this type is shown in Fig. 43. It differs from the temperature controller represented in

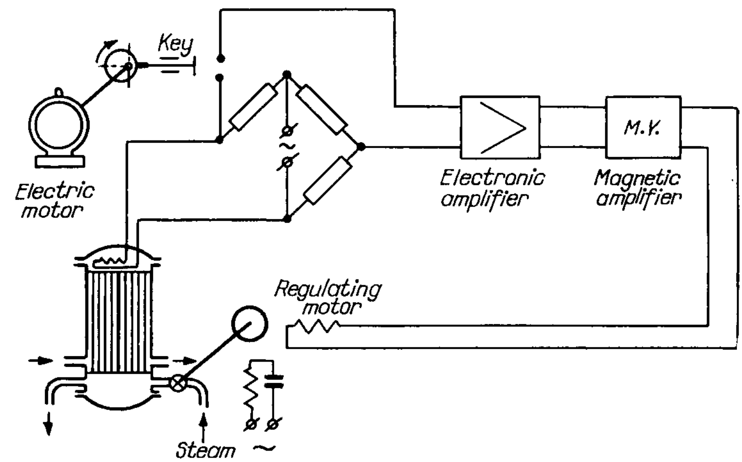


FIG. 43

Fig. 18 only by the presence of an interrupter (key) which is periodically actuated by a cam driven by an electric motor. Due to this interrupter in the sensor circuit, the temperature is measured and acts further on the circuit not continuously but only during the time intervals when the key is closed. In these intervals the controller acts exactly in the same way as the continuous controller (Fig. 18). In the periods when the key is open the control element of the object does not move.

Figure 44 shows a sensor with a chopper bar which is often used in discontinuous controllers in place of a key. The chopper is raised and lowered by the cam or cam gear which is rotated by an electric motor. When the chopper is raised, the sensor does not act on the controller and the control circuit is open. During this time the control

element does not move. When the chopper is lowered, the sensor pointer presses against one of the potentiometer arms and the circuit closes. A rectangular pulse is produced at the output of the sensor,

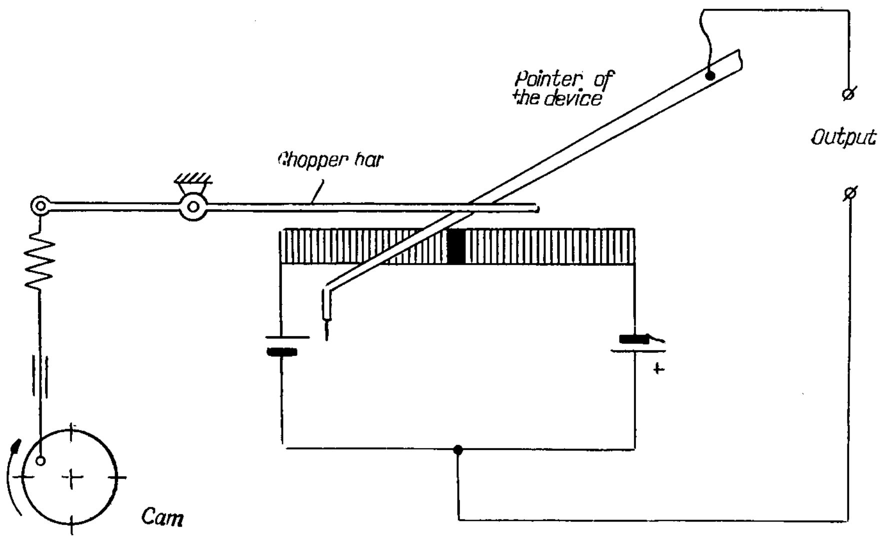


FIG. 44

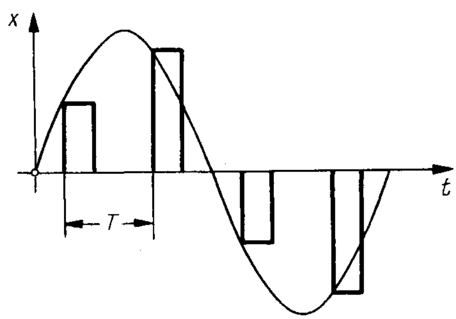


FIG. 45

its height depending on the position of the pointer at the moment when the chopper bar presses it to the contact.

Figure 45 shows the sequence of pulses obtained at the output of the sensor when the controlled quantity changes sinusoidally.

A different construction using a chopper bar is shown in Fig. 46. Here the chopper presses the pointer to one of two possible contacts,

which of the two being determined by the sign of the input. The height of the pulse produced at the sensor output when the pointer makes contact is constant, and the width of the pulse (the time of contact of the pointer), because of the taper on the bar, depends on the position of the pointer at the moment of contact. The sequences of pulses for a sinusoidal input signal in this case is represented in Fig. 47.

The theory of discontinuous control shows that discontinuous controllers have several advantages when compared with continuous

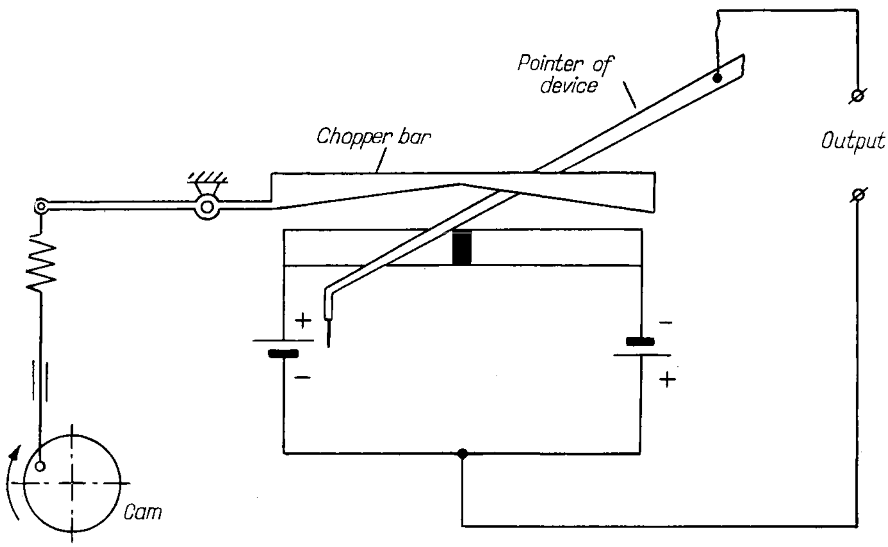


FIG. 46

controllers, in particular in the control of objects with large delays. But the use of discontinuous controllers in the automation of production is often explained not only by these, but also by purely technical advantages. When a chopper bar is used, ordinary low power pointer devices may be employed, without any loss of accuracy, since measurement and setting are separated in time. During measurement the device is cut off from the controller and no additional load from the side of the subsequent elements of the controller is transmitted to its moving system.

Also in discontinuous control the controller is disconnected from the object for a large part of the time. This enables one to use the

same controller for several (sometimes several dozen) parameters in turn (every cycle or according to a given curve). For this purpose the controller is connected in turn during one or several pulses to each of the sensors and to the corresponding regulating unit and moves the latter in correspondence with the indications of the sensor.

10. Extremal Controllers

Let the value of the controlled parameter be y and the position of the control element be x. In some cases the static characteristic

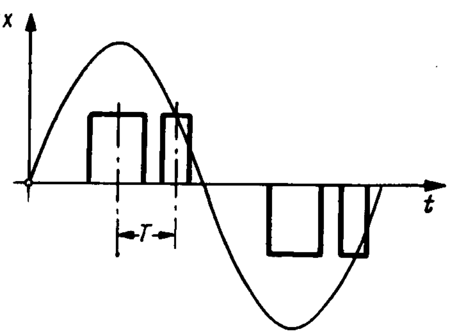


FIG. 47

of the object, $y = f(x)$, has an extremum (maximum or minimum). Depending on various factors the function $f(x)$ will change and the point corresponding to the extremum will move in the xy-plane (Fig. 48). It is often required to maintain y at the value corresponding to the extremum of the function $y = f(x)$.

For example, let us consider the problem of maintaining the highest temperature in the flame of an oil-burner.

The sources of the oxygen used for the fuel combustion are varied: primary air, secondary air, oxygen blowing, "unorganized air" (the seeping through thin material) and so on.

Let us suppose that the burner is given a constant quantity of fuel and that there is a damper which can, for example, change the supply of secondary air.

If the quality of the fuel and the quantity of oxygen supplied by the other sources were constant, then in order to maintain the

maximum flame temperature a definite quantity of secondary air would be required. An ordinary controller could maintain it. But if the oxygen received from the other sources changes then in order to maintain the maximum temperature the quantity of secondary air must also be changed. This problem can be solved by an ordinary controller, but it is then necessary to measure the supply of oxygen from all the sources, and the quality of the fuel, and to reset the setter of the secondary air controller so that the rate of flow of secondary air maintained by it would always secure the maximum temperature

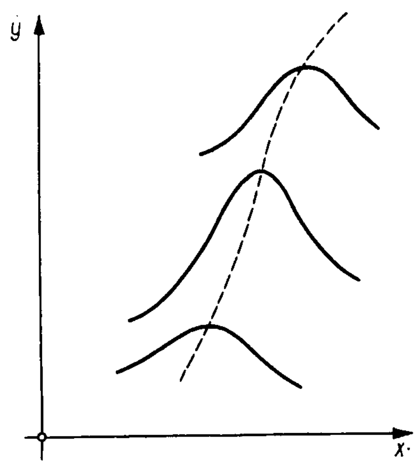


FIG. 48

of the flame. This is complicated, and often not possible; it is difficult, for example, to measure the accidental inflow of air through various thin surfaces in a stove. This problem can, however, be solved by a controller which is such that only the controlled quantity is measured (in this example, only the flame temperature) and the controller itself continuously "searches" for the position of the control element giving to the measured quantity its maximum or minimum value. A controller of this kind is called an *extremal controller*. At present, extremal controllers are only just beginning to be used in automation. But their advantages are indisputable, and in the near future they will be widely used.

The simplest scheme of an extremal controller is shown in Fig. 49. The value of the controlled quantity, y, which is measured by the sensor, is transmitted to the differentiating unit. Its output signal is

proportional to the derivative $\frac{\mathrm{d}y}{\mathrm{d}t}$. Later in the circuit there is a divider, which receives one signal proportional to $\frac{\mathrm{d}y}{\mathrm{d}t}$ and another proportional to the rate of change of the control element coordinate $\frac{\mathrm{d}x}{\mathrm{d}t}$. The output from the divider is therefore proportional to their ratio $\frac{\mathrm{d}y}{\mathrm{d}x}$.

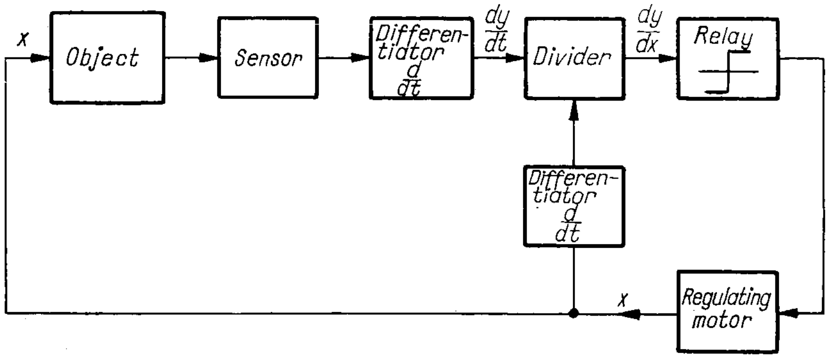


FIG. 49

When this derivative changes in sign, a relay reverses the regulating motor, i.e. changes the direction of change of the control element. We can consider the unit as a whole as an ordinary controller maintaining the quantity $\frac{\mathrm{d}y}{\mathrm{d}x}$ at a value $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$. The sensor, reacting to the deviation in y, can be considered, together with the differentiating unit and the divider, as a sensor reacting to the controlled parameter $\frac{\mathrm{d}y}{\mathrm{d}x}$. But the value of this parameter becomes indeterminate if at the same time $\frac{\mathrm{d}x}{\mathrm{d}t} = 0$ and $\frac{\mathrm{d}y}{\mathrm{d}t} = 0$. Hence such a controller can operate only if the control element does not stop when the maximum of $y = f(x)$ is attained. This is ensured by the use of a two-position relay, i.e. by constructing the controller according to a two-position (oscillatory) scheme.

The basic disadvantage of a controller of this type lies in the fact that it is very sensitive to noise. In the presence of noise the characteristic $y = f(x)$ has many small extrema (Fig. 50). The controller

has to be considerably more complicated if it is to distinguish the required (true) extremum from the bogus extrema.

A different scheme for an extremal controller is represented in Fig. 51. Both the regulating motor of the controller and the constant frequency oscillations produced by a special oscillator act on the control element.

Let us first suppose that the control element produces only oscillatory movements. The oscillations of the controlled quantity y at the output of the object are produced with the same frequency, but their

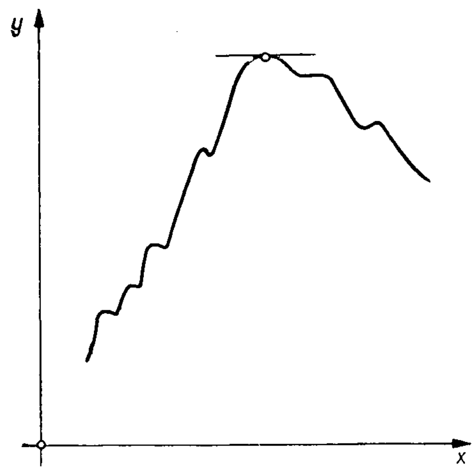


FIG. 50

phase depends on the sign of $\frac{dy}{dx}$ from the characteristic $y = f(x)$. When the extremum is crossed, the sign of the phase changes.

In the scheme of Fig. 51, the value of y is transmitted to the input of a phase-sensor, which reacts to the sign of the phase of the oscillator signal depending on its value, switches the relay which reverses the regulating motor.

As a result, the control element takes part simultaneously in two motions: translatory and oscillatory. The oscillatory motion is used to obtain a signal to determine whether the required extremum has been attained. When the extremum is passed the direction of the translation of the control element is changed.

Another method for constructing an extremal controller is shown in the scheme of Fig. 52.

The controlled quantity is transmitted to the input of the memory and of the comparing unit. In the memory unit the greatest (or least) value attained by the controlled quantity is stored. In the comparing unit the value of the controlled quantity which is stored in the memory unit is compared with the current value. At the moment that the difference between them reaches a given magnitude (for example, 0·5 or 1 per cent of the value which is stored in the memory), an output

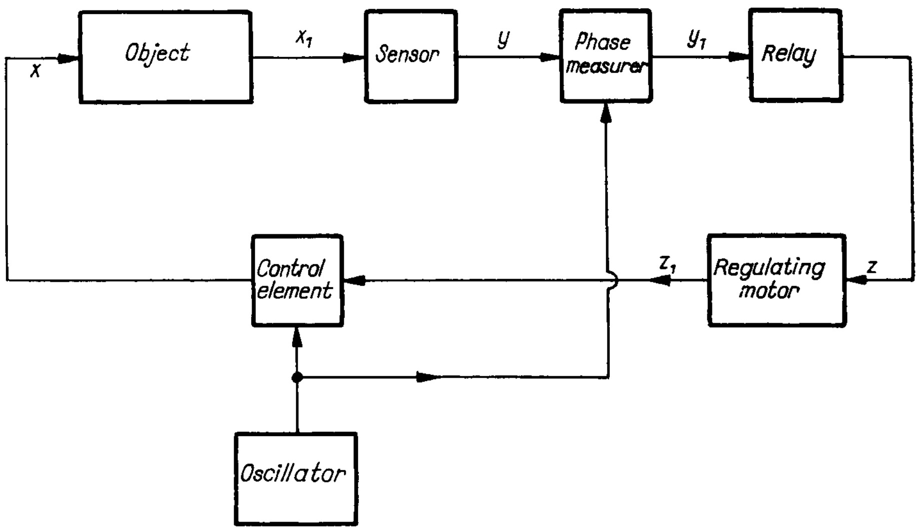


FIG. 51

signal is obtained from the comparing unit, which fulfils two functions: it switches the relay which reverses the direction of motion of the control element and it clears the store of the memory unit.

The scheme of a pneumatic extremal controller, designed at the Institute of Automatics and Telemechanics of the Academy of Sciences of the U. S. S. R., is shown in Fig. 53.

The sensor produces a pneumatic signal P_y, proportional to the measured quantity. This signal is transmitted to the chamber A and also to the chamber Z of the memory block and to the chamber B of the comparison block. The pressure from the memory chamber in its turn reaches the chamber C of the memory block and the chamber D of the comparison block. When P_y increases, the air flows from chamber A through the memory chamber Z to C. The pressure in A,

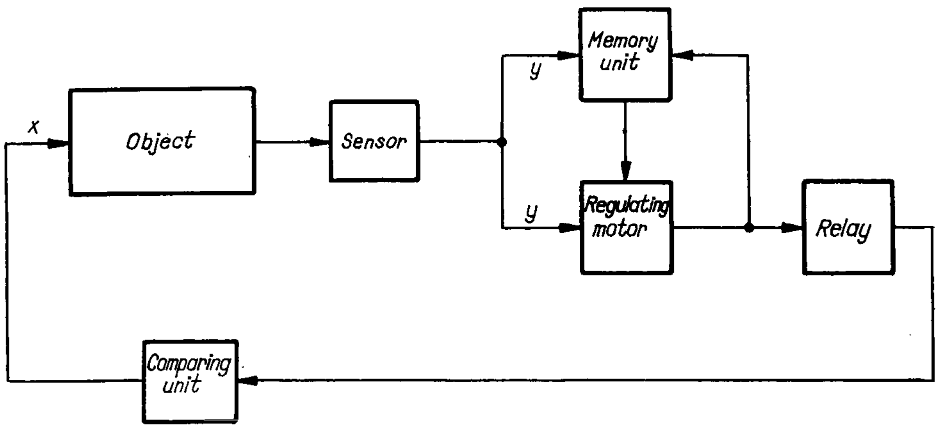


FIG. 52

due to hydraulic losses, is slightly more than that in C, and the membrane unit of the memory block is pressed downwards. The nozzle C_1 is closed by the valve, the pressure in chamber E becomes equal to the total supply pressure and the lower membrane of the memory block, pressing downwards on the valve K, opens the access of air from the transmitter across the memory chamber Z to C.

At the moment when P_y attains its maximum and begins to decrease, the direction of flow is changed: the air from Z must now

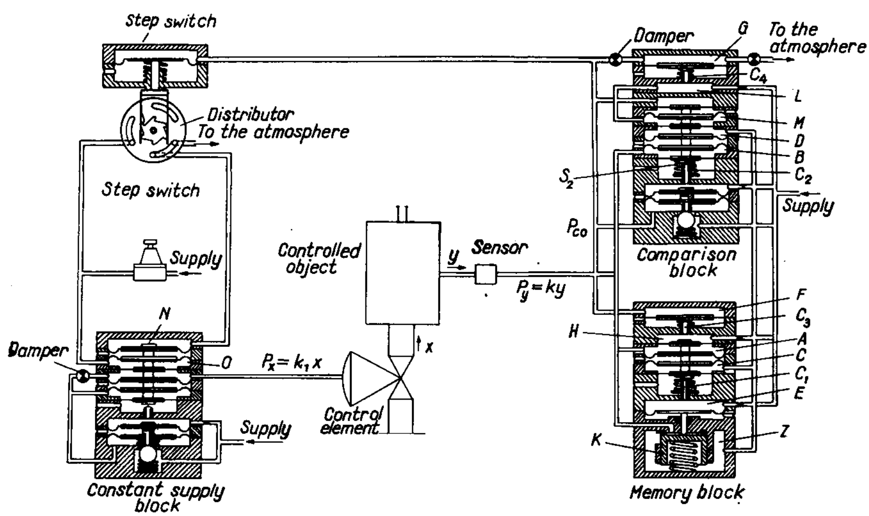


FIG. 53

flow into A. But the pressure in B will then become larger than that in A, the membrane is pressed upwards, the nozzle C_1 is opened and the pressure in E falls. Because of this the valve K is slammed shut and the maximum pressure is attained in Z and, of course, in C and D.

In the comparison block, this fixed pressure $P_{y\max}$ is compared with the instantaneous pressure P_y which is transmitted to the chamber B. When the difference between these pressures reaches a given value δ (set by the tension in the spring S_2) the nozzle C_2 closes and the pressure at the output of the comparison block, P_{CO}, increases sharply.

This pressure P_{CO} is transmitted to F in the memory block, to the membrane of the pneumatic step switch and across the choke which ensures a delay in the signal, to the chamber G of the comparison unit. Under the action of the given signal, the step switch turns the ratchet-wheel and the plane air distributor joined to it, by one tooth.

The increase in pressure in F causes the membrane to sag downwards and the nozzle C_3 to close. The pressure in H increases, the membrane unit sags downwards, the pressure in E increases and the valve K opens. Because of this the pressure in the memory chamber Z becomes equal to P_y, i.e. the "memory" of the memory block is cleared.

After some delay, the pressure in G increases. This switches the nozzle C_4 and causes an increase in the pressure in chambers L and M. The membrane unit of the comparison block presses upwards and the signal P_{CO} is removed.

As a result, the memory and comparison blocks return to their initial position and are again ready for operation. But now the step switch distributor has been turned on by one step.

At each turn of the distributor the pressure is alternately transmitted to the chambers N and O. Depending on this, the constant pressure block directs the air to the control element membrane, and from it to the atmosphere, maintaining a constant drop at the throttle, through which the air flows. As a result for each operation of the step switch the direction of motion of the control element, which moves with a constant velocity, is changed.

Thus the control element continuously oscillates about the value corresponding to the maximum of the controlled quantity y.

Let us now turn to a general consideration of extremal controllers. Two ways of using extremal controllers in the automation of production can be indicated.

When the quantity y, having an extremum for some position of the control element, is measured directly, its value is immediately transmitted to the input of the extremal controller (Fig. 54). In a number of cases the quantity y cannot be measured directly (for example when y is the efficiency of an assembly or the specific rate

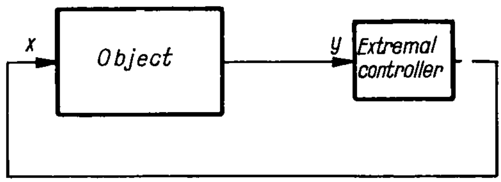


FIG. 54

of flow of fuel) but it can be calculated from parameter measurements. In this case (Fig. 55) the measured parameters are transmitted to the input of a computing device in which the value of the parameter y whose extremum must be maintained is computed. The extremal controller reacts to this parameter calculated in the computing device.

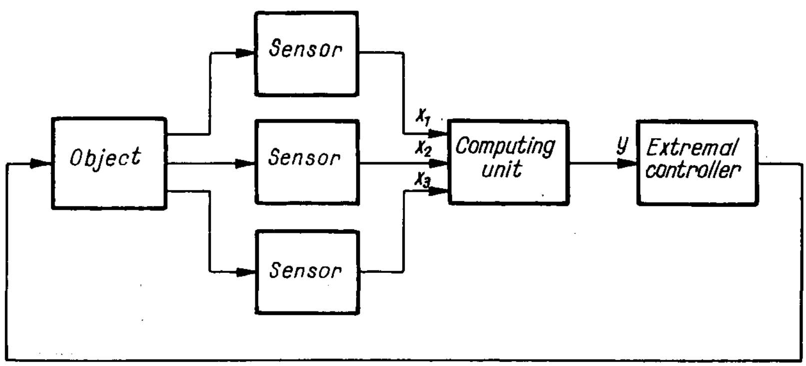


FIG. 55

The principles described above can be applied also in the construction of an extremal controller for several quantities.

Let us suppose, for example, that the controlled object has several (say two) control elements. Let their coordinates be, respectively, x_1 and x_2. Let the controlled parameter $y = F(x_1, x_2)$ have a maximum (or minimum) for some values x_{10} and x_{20} (Fig. 56). The

surface $y = F(x_1, x_2)$ changes continuously, the values x_{10} and x_{20} corresponding to the extremum change, and the control elements must always be set in the positions x_{10} and x_{20}.

To obtain this, the usual extremal controllers may be used, if with their help we move the control elements alternately.* Let the value $x_1 = \bar{x}_1$ be fixed and let the extremal controller control the second control element. It then operates according to the characteristic

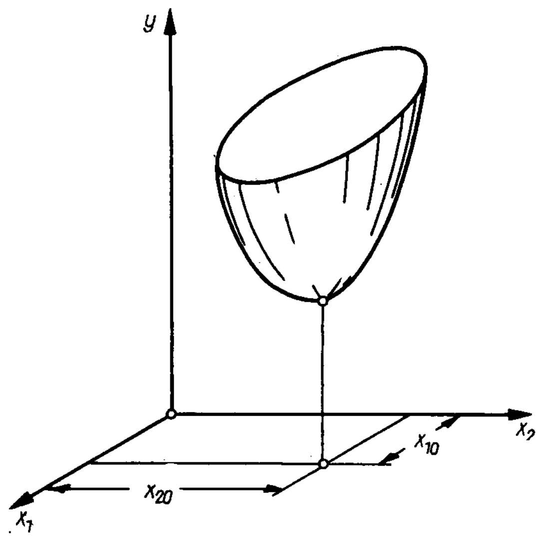


Fig. 56

which is obtained if we intersect the surface $y = F(x_1, x_2)$ by the plane $x_1 = \bar{x}_1$, and bring y to the minimum point in the section. We then fix $x_2 = \bar{x}_2$ and the same extremal controller controls the first control element and so on, until we obtain the absolute extremum (Fig. 57).

Controllers of this kind allow us to construct a system of control by objects with several controlled quantities, by a method different in principle from that used in ordinary controllers, even in those cases when the problem is not to maintain the extremum but is the usual problem of stabilizing the desired values of the controlled

* Other methods of solving this problem are also possible. We can, for example, simultaneously move all the control elements so that the point drawn moves along the surface to the minimum along the curve of shortest descent. But such a construction is more complicated.

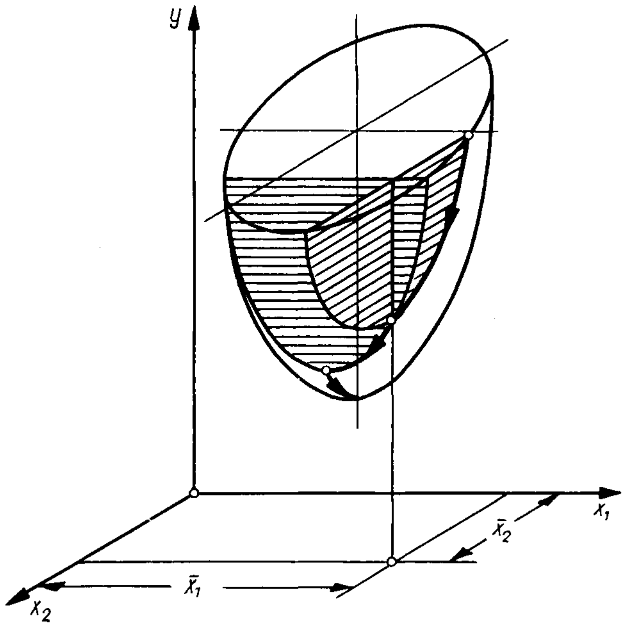


FIG. 57

quantities. In order to do this, the controlled quantities x_1, x_2, ..., are transmitted to the input of the computing device (Fig. 58) which produces at its output the *free parameter*, some function of these quantities $y = F(x_1, x_2, \ldots,)$. This function is chosen so that it has a

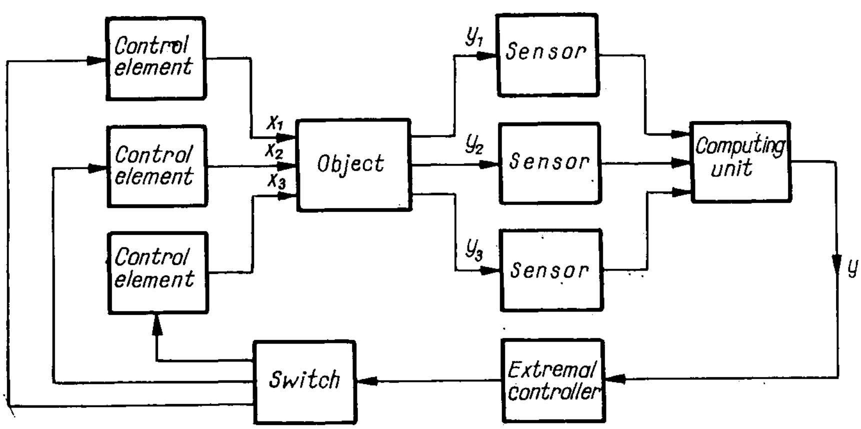


FIG. 58

sharp minimum (or maximum) for the values of x_{10}, x_{20}, . . , which it is necessary to maintain. Then the free parameter is transmitted to the input of the extremal controller which moves all the control elements until the extremum value of the free parameter y is obtained, i.e. until $x_1, x_2, \ldots$ take these required values. Of course, the positions to which the extremal controller moves the control elements will depend on the external disturbances.

11. Individual, Specialized and Universal Controllers. Aggregate Systems

The calculation and construction of controllers depends essentially on the purpose and peculiarities of their production.

Controllers which are intended for the control of a particular object are called *individual.* Usually such controllers are supplied complete with the object and are an integral part of it. Such for example are the controllers of powerful turbines, jet engines, and so on.

During the design stage of such controllers, the peculiarities of the controlled object are considered as data, and the purpose of calculation is to ensure the conditions which are laid down by the technical data of the control process during the operation of this concrete object.

Specialized controllers are intended to control a particular parameter in various objects. As examples we may take almost all direct action controllers, indirect action pressure controllers, many temperature controllers, and so on. The sensor and in part the regulating unit of such controllers are an integral part of them and cannot be altered. But it is not known beforehand, during construction and manufacture, where, with what objects and under what conditions the controller will be operated.

Universal controllers are produced by specialized controller-constructing enterprises to control the most varied parameters and objects. Universal controllers are made so that various sensors and very varied regulating units can be adapted to them. Not only during construction and manufacture, but also at the time of sale, it is not known where the controller will operate, under what conditions, or even what parameters are to be controlled. This being so, there is no point in designing the control process, and the calculation of the controller is carried out separately with regard to the sensor or the object.

The controller is designed as a separate unit intended for the transformation of the input signal corresponding to the required law of the control effect. Great attention is paid, therefore, to securing a wide tuning range. Thus, in a number of universal controllers, the effectiveness of the feedbacks may change by a hundred or a thousand times, and the time of a floating component may change from a few seconds to several hours.

In recent years, as a result of the wide development in the automation of production, everywhere the production of universal controllers is being replaced by the production of even less specialized devices: aggregate systems of automatic control. Such systems consist of separate units, intended to fulfil the simplest operations on a signal, i.e. amplification, compounding signals, differentiation, integration, and so on. Any control system consists of such standard units, so that very varied automatic schemes may be put together from a small selection of units. The analysis of each unit must be done separately, depending on its working conditions.

More and more frequently, conditions arise where the closed control circuit is completed by the consumer in his own way, and the analysis of the control system by the manufacturer loses its value. In addition, when there are similar tunings over a wide range, the consumer seldom resorts to an analysis of the closed control circuit. He requires something else from control theory: a convenient, simple method of discovering optimal tuning. The discovery of such methods has until now been rather neglected. Nevertheless, without a knowledge of control theory it is not possible to construct and adjust controllers, or to complete modern control systems. The theory enables us to understand the causes of self-excitation of control systems, to design a correct programme of experimental running of the controller, and of debugging the control system. For individual and for some types of specialized controller, control theory enables us to design the controller and to foresee during its construction everything necessary for its prompt running.

The remaining chapters of this book are devoted to the methods of control theory and to the elucidation of the possibilities offered by it. Only the theory of continuous control is dealt with. A knowledge of it will simplify the study of special subjects in control theory, the theory of discontinuous control and of the extremal controller, for the reader.

CHAPTER II

THE CONSTRUCTION OF A LINEAR MODEL CONTROL SYSTEM AND THE INITIAL MATERIAL FOR ITS ANALYSIS

THE linear analysis techniques of automatic control enable us to solve control problems to the first approximation, that is, to determine what kind of control process in a given system will follow a given disturbance, or, conversely, to find those controller parameters in the given circuit for which a process satisfying the given conditions takes place. However, problems of this kind can be completely solved only for some types of system, namely those defined by linear differential equations. Usually control systems are defined by more complicated, non-linear, differential equations, and therefore the initial steps of the analysis consists in replacing such equations by appropriately chosen linear equations, and in using these to study the processes in the given system. In some cases, such a method succeeds in elucidating the character of the processes in the actual system or even in completely solving them, while in other cases it assists the study of the processes only for sufficiently small disturbances. However, it is usual in any case to begin the solution with such a linear analysis. If the linear analysis is inadequate, then it becomes necessary to examine considerably more difficult, non-linear, problems in order to solve the system. Methods for their solution have been little developed, and if we need to investigate not only the steady state but also the transients of the processes, then we must appeal to graphical or numerical methods for the solution of the non-linear differential equations of the control process. In doing this, we have normally to set the parameters "at random", until we have succeeded in choosing those which ensure processes satisfying the technical data.

From this brief description of the state of the theory of control and its resources, it is evident that any research into the control system requires, first of all, the derivation of the differential equations defining the control process and the ability to obtain from these equations their corresponding linearized form, which will be suitable for investigation by the methods of the linear theory of control.

We must find first the static solution of the control system rather than proceed to the derivation of the equations of motion, since we will succeed in finding the steady values reached by all the coordinates of the system for any investigated conditions with the static solution. These steady values assumed by all the coordinates must be known for the derivation of the equations of motion in non-linear systems.

It is convenient, in deriving the differential equations, to dismember the system into parts, and according to the laws of uniformity, to derive the equations of motion for each part of the system separately.

Sections 1—3 of the present chapter are devoted to all the questions concerning the derivation of the equations of motion.

When the equations of all parts of the system have been obtained, we then find their corresponding linear equations. From these is found the transfer function of the given system (these terms will be explained in more detail below), which is also used as initial material for all further linear calculations. Sections 4—7 are devoted to questions of this kind.

The methods of the linear theory of control can be divided into two groups. To the first belong those methods which study the process directly from the linearized equations. For these the presentation of the initial material in the form of the transfer function is very convenient. But other methods, comprising the second group, are also widely used. These start from the presentation of the properties of the linearized system in the form of special graphs, frequency characteristics, which are easy to construct from the transfer function. But in a number of cases, where the actual system differs little from its linear approximation, the frequency characteristics can be obtained experimentally; this constitutes one of the important advantages of the frequency response method, since in such cases one can sometimes totally avoid the derivation and linearization of the equations of motion. The initial material for frequency response methods consists

of frequency response characteristics either constructed from experiments or obtained from the linearized equations. Everything connected with obtaining it is summarized in Section 8 of this chapter.

1. The Dismemberment of the System into its Elements

Corresponding to the accepted terminology of mechanics, by the *number of degrees of freedom* of a system of automatic control is meant the least number of independent quantities whose values completely determine the state of the system at any instant. The quantities satisfying these conditions are called the *generalized coordinates of the system*. These quantities can be chosen arbitrarily and can have any dimension.

If we were interested in all the processes taking place in a system during automatic control the number of degrees of freedom would be extraodinarily large or even, if the distribution of parameters were taken into account, infinite. However, to the first approximation, we can distinguish between those generalized coordinates which basically affect the course of the control process and those which have very little influence on it, and in this way the number of degrees of freedom taken into account can be reduced.

By comparing the time required to set up a coordinate with the expected or required tempo of the control process, we possess a criterion for selecting the basic generalized coordinates. If for an intermittent change in some coordinate A the new value of the coordinate B which follows it (in the action circuit) is set up in a time incommensurably less* than the expected or required time of the control process, then we need not introduce the coordinate B into the problem, but consider its value to be "due" to the value of A.

Let us consider some examples of selecting generalized coordinates.

In the direct automatic control assembly for the speed control of a diesel engine, shown in Fig. 59, we can restrict ourselves to two generalized coordinates and, therefore, we can think of the whole system as having two degrees of freedom. As one generalized coordinate we may take the mean angular velocity per revolution of the diesel crankshaft, and as the other the position of the controller clutch

* Usually by some tens of times.

or else the position of the fuel pump lath. Being restricted to these two coordinates we can completely disregard the irregularity of the speed of the crankshaft (the change of its angular velocity over one revolution), the wave processes in the fuel pipes and in the rods connecting the controller clutch and the fuel pump lath, and so on, considering that these processes take place incommensurably faster than the control process itself.

In a single-stage reducer also (a direct action vacuum controller, Fig. 60) it is natural to limit consideration to only two generalized co-

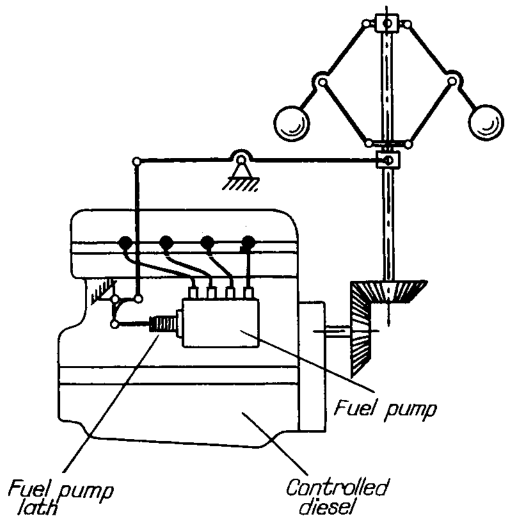


FIG. 59

ordinates. One is the pressure under the diaphragm, and the second is the position of the diaphragm (or of the reducer valve). It is then assumed that the pressure at the reducer input does not vary and that the wave processes in the chamber under the diaphragm can be neglected, i.e. it is considered that a uniform pressure is set up practically instantaneously at all points of this chamber. Similarly in a two-stage gas reducer (Fig. 61) four degrees of freedom are taken into account: we can take as generalized coordinates, for example, the deflections of both membranes, the pressure between the stages and the low pressure at the reducer output.

When using an indirect controller with an hydraulic jet amplifier to control the rate of flow, it is necessary to take three degrees of

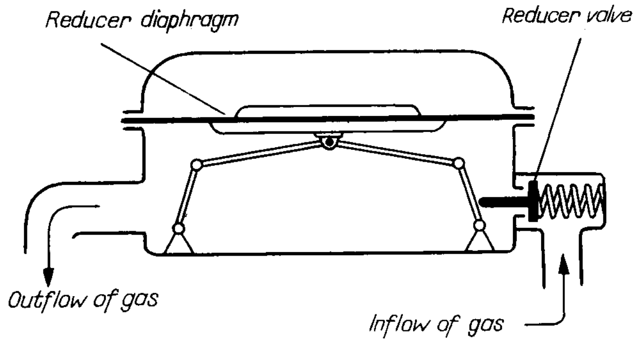


FIG. 60

freedom into consideration (the generalized coordinates being the rate of flow, the movement of the jet pipe and of the piston). Often it is possible to treat the rate of flow as being "due" to the movement of the control element, and to reduce the number of considered co-ordinates to two.

In a system for automatic air pressure control using the controller 04-MG (its circuit was shown in Fig. 27) seven degrees of freedom have to be taken into account. The generalized coordinates can be the controlled pressure, the angle turned through by the end of the helical spring, the pressure in the primary relay chamber, the tension in the bellows, the pressure in the secondary relay chamber, the position of the valve in the control element and the displacement of the feedback ever.

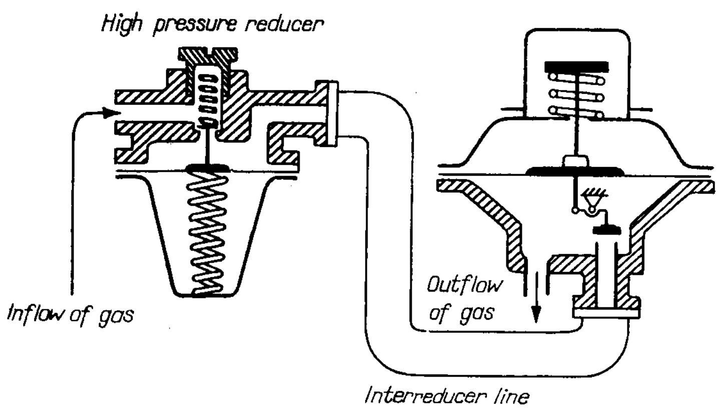


FIG. 61

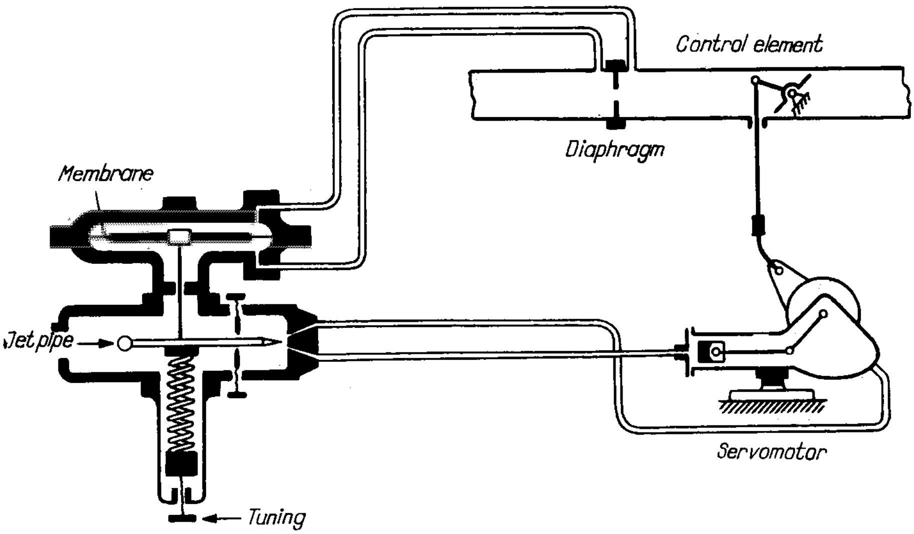


FIG. 62

If the control process as a whole takes place slowly (for example if the control time lasts several minutes) then the number of degrees of freedom considered can be reduced. We may, for example, consider in this case the pressure in the primary relay chamber and the tension in the bellaws as being instantaneously set up, i.e. as being precisely "due" to the angle turned through by the end of the helical spring.

As an example of this, let us consider the control of the angular velocity of a direct current electric motor with independent excitation.

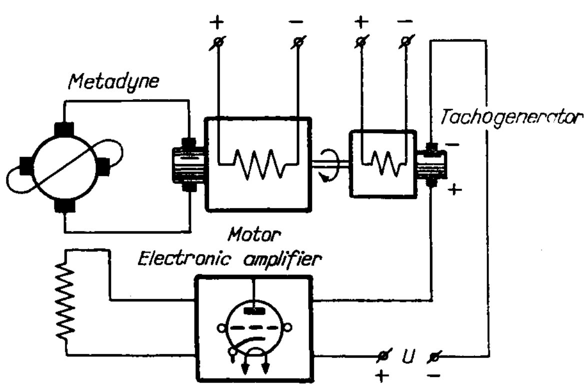


FIG. 63

The control is effected by the influence of the motor armature on the potential (Fig. 63). The angular velocity is measured in this example by a tachogenerator, the e.m.f. of which is compared with a standard potential U. The difference between them after amplification by an electronic amplifier, is transmitted to the excitation coil of a metadyne whose output brushes supply the armature of the controlled motor.

If the electronic amplifier is considered as having no inertia, then the system has five degrees of freedom. The generalized coordinates may be taken as the controlled speed of the motor, the voltage at the electronic amplifier output, the current in the excitation coil of the metadyne, the current in its short-circuited circuit and, finally, the armature network of the controlled motor.

We note the fact that in control systems the action between two units of the system often proves to be one-sided, one unit acting upon the next without meeting any noticeable opposition.

Thus, for example, in the pneumatic controller 04-MG (Fig. 27) the turn of the end of the helical spring is determined by the deviation in the stage of the controlled member, but this state does not change with the addition of the helical spring; the movement of the end of the spring causes a change of pressure in the primary relay chamber, but this change of pressure has almost no influence on the position of the end of the spring; the pressure in the primary relay chamber determines the pressure of the bellows, but the pressure in this chamber is hardly affected by the displacement of the bellows, and so on.

Henceforth any unit which allows action in one direction only, i.e. which, apprehending an action from the preceding part of the system, offers it negligible opposition, is called *directional*.

If at any point of the circuit we select some directional unit and we break the action circuit in front of it, i.e. cause the actions of the preceding part of the system to influence this directional assembly in no way, then the system as a whole will be called *open*.

We can pick out that element in the selected directional unit on which the action was carried out in the earlier, closed, system. It is called the *input of the open system*, and the coordinate determining its state is the *input coordinate of the open system*. The element in the last part of the open system which carried out the action on the given directional element when the system was closed is called the *output of the open system*, and the coordinate characterizing the state of this element, its *output coordinate*.

Thus, for example, in the case of the direct control of a diesel we can open the system by disconnecting the pump lath from the controller clutch. The pump lath will then be the input of the open system and the position of the lath will be the input coordinate; the controller clutch will be the output of the open system and the position of the clutch its output coordinate. We can open the same system by disconnecting the engine flywheel from the drive gear of the controller shaft. Then this gear will be the input of the open system, its angular velocity the input coordinate, and the engine flywheel will be the output of the open system and its angular velocity the output coordinate.

In an open system we can again select some directional unit and similarly break the action circuit in front of it. The open system is then divided into two parts. If either of these parts contains more than one directional unit it can again be divided into two parts, and so on.

As a result the system will be divided into sections each of which contains only one directional unit. Such a section of the system is called an element of it. Just as for the open system, each element has its "input" and "output" and corresponding "input" and "output" generalized coordinates. The input coordinate of each element is the output coordinate of the element preceding it.

Thus, for example, the system of automatic control of a rate of flow (Fig. 62) can be dismembered into three elements. The measured rate of flow is the input coordinate for the measurer and the output coordinate for the controlled member; the position of the end of the jet tube is the input coordinate for the servo-motor and the output for the measurer; finally, the position of the controlling valve is the input coordinate for the controlled member and the output for the servomotor.

The theory of automatic control enables us to judge the properties of a system from the properties of its constituent elements.

All generalized coordinates will henceforth be denoted, regardless of their dimensions, by X. We will denote the input coordinate by X_{in}, the output coordinate by X_{out}, and the number of a co-ordinate will be indicated by a suffix (for example $X_{\text{in}1}$, $X_{\text{out}2}$, etc.).

As the zero reading of each coordinate we choose the value which it has for some steady operating condition of the system.

The direction of reading the generalized coordinates can be selected arbitrarily, but it is convenient to keep to the following rule

of signs: the controlled coordinate (the coordinate X_1) is positive if its value is larger than that which is selected as the reading origin; the coordinate of the sensor X_2 is positive if an increase in X_2 corresponds to an increase in X_1, and so on down the action chain.

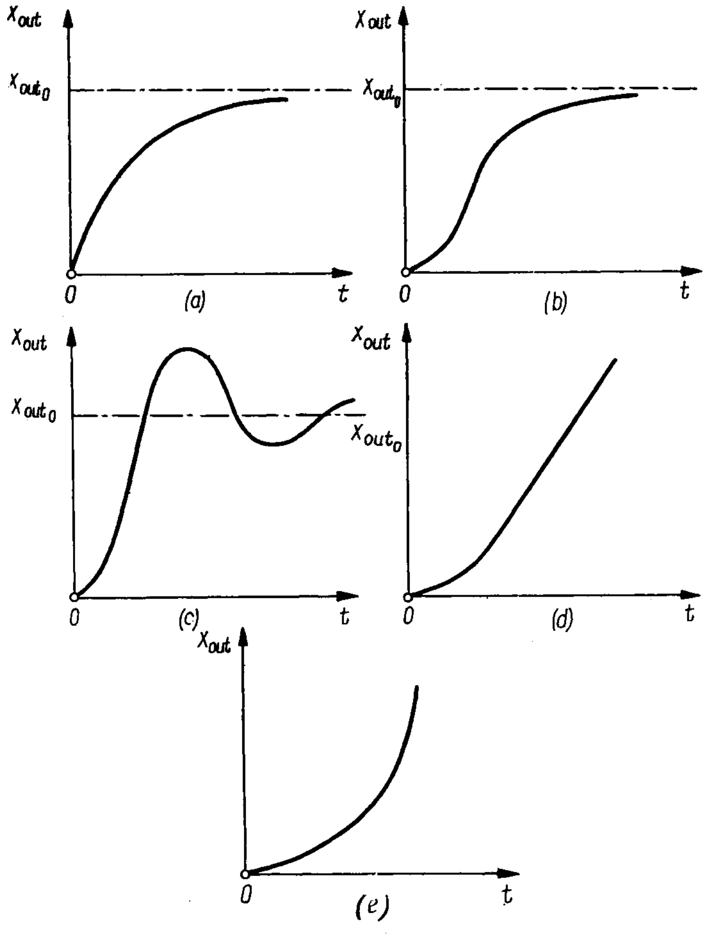


FIG. 64

With this sign rule, the output coordinate of an element will be the input coordinate of the next element not only in magnitude but also in sign.

The *time characteristic* of an element is the curve which shows the change of its output coordinate as a function of time resulting

from its input coordinate instantaneously acquiring a new value which is not subsequently altered.

In order to determine the time characteristic using any recording device (for example an oscillograph with corresponding monitors) we make a simultaneous recording of the change in the input and output coordinates X_{in} and X_{out} of the element. Having switched on the recording device, we abruptly change the value of the input coordinate from some value $X_{\text{in}1}$ to $X_{\text{in}2}$ and during subsequent recording this value $X_{\text{in}2}$ remains constant. On the oscillograph there will be drawn two lines which, apart from time delays, will correspond to the change in X_{in} and X_{out} during the experiment.

Figure 64 shows examples of typical time characteristics. An element whose output coordinate tends over the course of time to a new fixed value is called *static* (Fig. 64*a*, *b* and *c*). If its output coordinate does not tend to a new fixed value and over the course of time a constant rate of change of the output coordinate is set up, then the element is *astatic* (Fig. 64*d*). If, finally, the rate of change of the output coordinate (and also its acceleration and higher derivatives) increase without limit, the element is said to be *unstable* (Fig. 64*e*).

2. Static Characteristics of Elements and of Systems

We restrict ourselves for the present to a consideration of static elements. Let us take a series of time characteristics, varying the value of the input coordinate $X_{\text{in}\,0}$ supplied to the element. We denote by $X_{\text{out}\,0}$ the value of the output coordinate which is ultimately attained when $X_{\text{in}} = X_{\text{in}\,0}$.

As abscissa we take the set-up value of the input coordinate $X_{\text{in}\,0}$ of any element, and as ordinate the set-up value of the output coordinate $X_{\text{out}\,0}$ of this element. If we ignore insensitivity, then for each element of the system we can construct its *static characteristic* in terms of the coordinates $X_{\text{in}\,0}$, $X_{\text{out}\,0}$ (Fig. 65).

In the case when some external action, as well as the action of the preceding element, acts upon the element, a family of static characteristics can be constructed for the element, each characteristic corresponding to a certain value of this external action.

Thus, for example, the controlled object possesses a family of static characteristics, each of which corresponds to a certain value of the load X_L on the object (Fig. 66). A sensor containing a setter

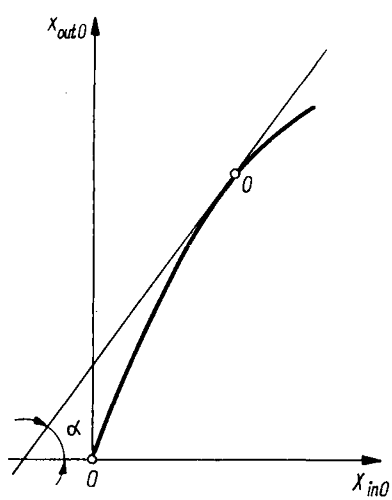


FIG. 65

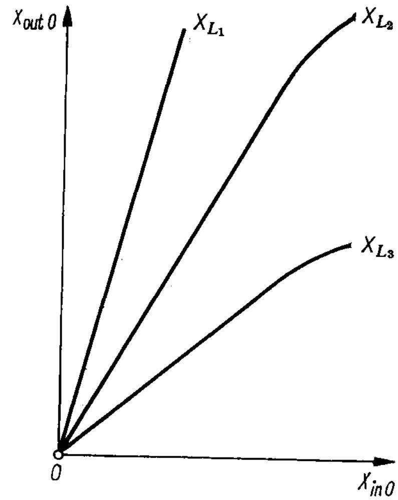


FIG. 66

also possesses a family of static characteristics, each of which corresponds to a certain position of the setter.

For static elements, the tangent at any point of the static characteristic has a positive slope:

$$\frac{dX_{\text{out }0}}{dX_{\text{in }0}} > 0.$$

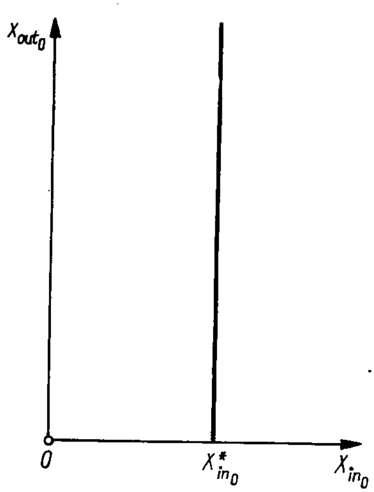


FIG. 67

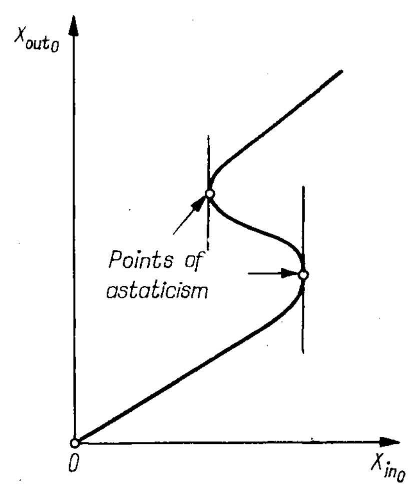


FIG. 68

The value of $\dfrac{dX_{out0}}{dX_{in0}}$ at some point 0 of the static characteristic is called the *coefficient of amplification* at that point and is denoted by k_0. From Fig. 65 it is clear that $k_0 = r \tan \alpha$, where r is the coefficient taking the scale of the $X_{out\ 0}$ and $X_{in\ 0}$ axes into account.

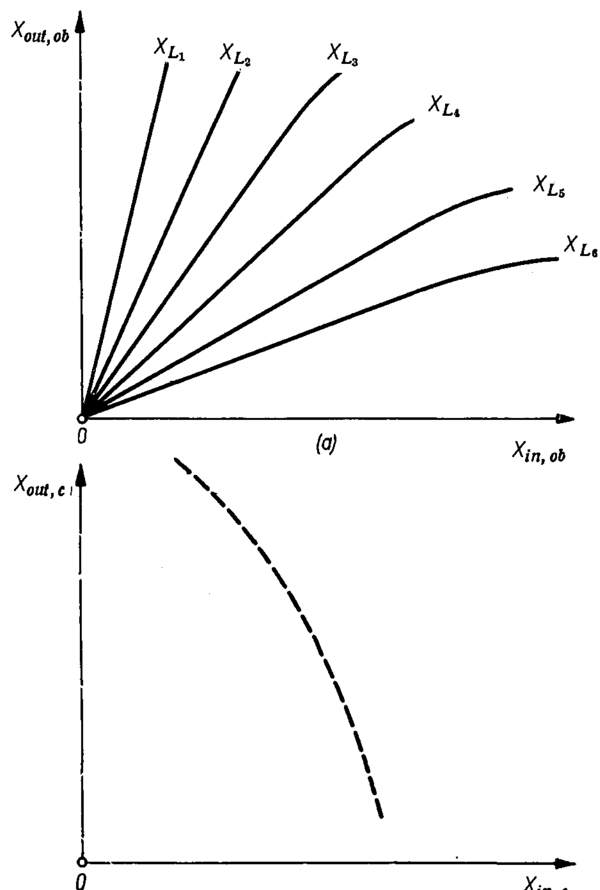


FIG. 69

An astatic element is, for some value of the input quantity $X_{in\ 0} = X^*_{in\ 0}$, in equilibrium for any value of the output quantity, but for other inputs $X_{in\ 0} \neq X^*_{in\ 0}$ has in general no equilibrium. We may thus think of the straight line parallel to the $X_{out\ 0}$ axis (Fig. 67) as the static characteristic of the astatic element.

If a vertical tangent can be drawn to the static characteristic of an element then the point of contact of this tangent is called a *point of astaticism* (Fig. 68).

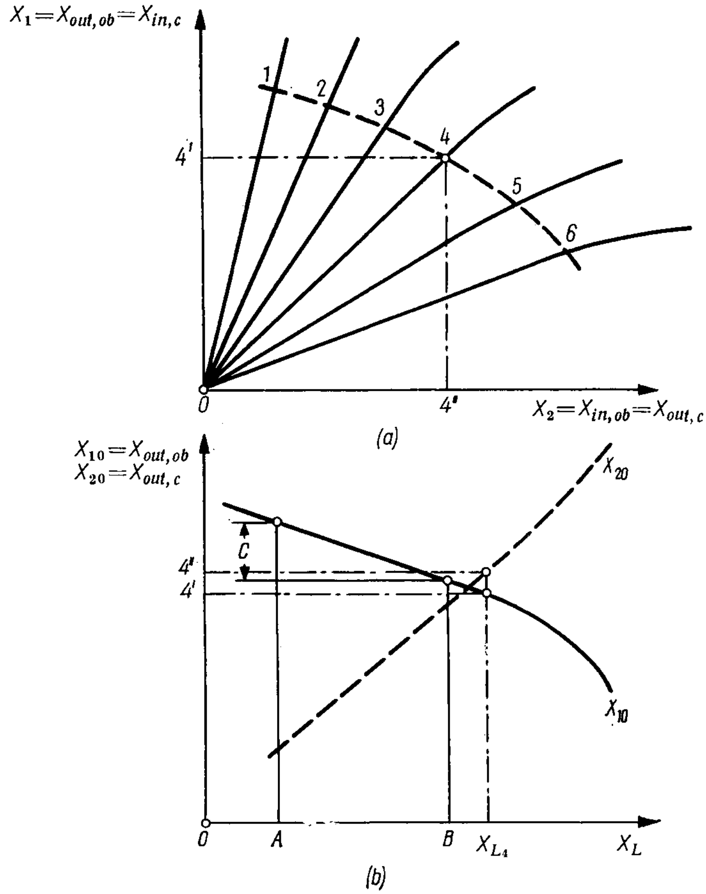


FIG. 70

k_0 is a measure of the statism of an element. The larger the value of k_0 the nearer to astaticism are the properties of the element.

The curves which define the change in the steady values of all the generalized coordinates of a closed control system for various values of the load or tuning are called the static characteristics corresponding to those coordinates. The static characteristic for the controlled co-

ordinate has a special significance. It is sometimes called the *static characteristic of the control system.*

The static characteristics of the control system can be constructed from the static characteristics of all the elements of the system.

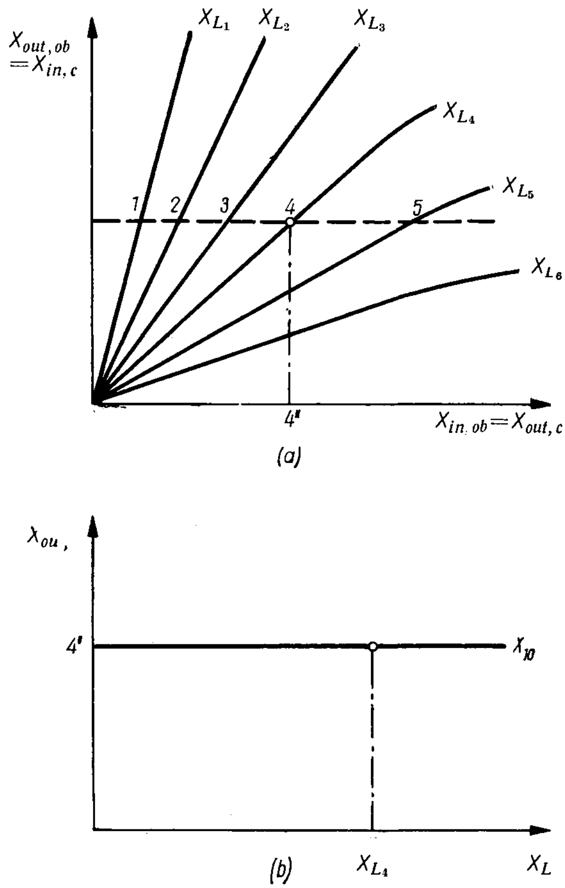


FIG. 71

We begin the construction of the static characteristics of the control system with a consideration of the process of direct control.

Let us suppose that the static characteristic of the controlled object for various values of the load X_L and the static characteristic of the controller are given (Fig. 69).

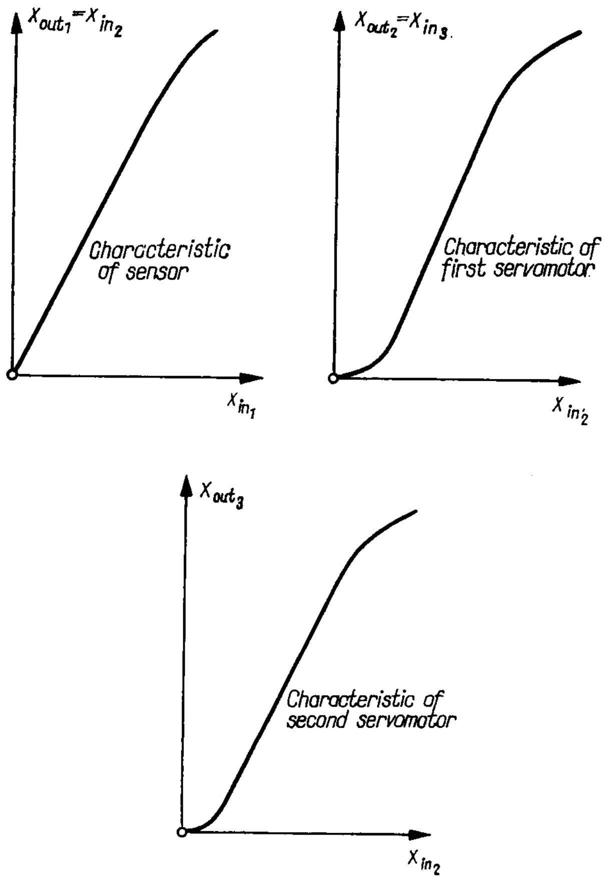


FIG. 72

Let us draw these static characteristics on one graph (Fig. 70), bearing in mind that the output co-ordinate of the object is the input coordinate of the controller and the output coordinate of the controller is the input co-ordinate of the controlled object, i. e.:

$$X_{\text{out, ob}} = X_{\text{in, cont}} = X_1,$$
$$X_{\text{out, cont}} = X_{\text{in, ob}} = X_2.$$

The points of intersection determine X_1 and X_2 for various loads in the steady conditions (Fig. 70a).

Using the points of intersection it is easy to contruct curves of change of X_{10} and X_{20} for a change in the load (Fig. 70*b*), i. e. the static characteristic of the control system.

If the control process is caused by a drop in the load from A to B, then X_{10} and X_{20} can be at once determined from Fig. 70*b*, both for load A and for load B.

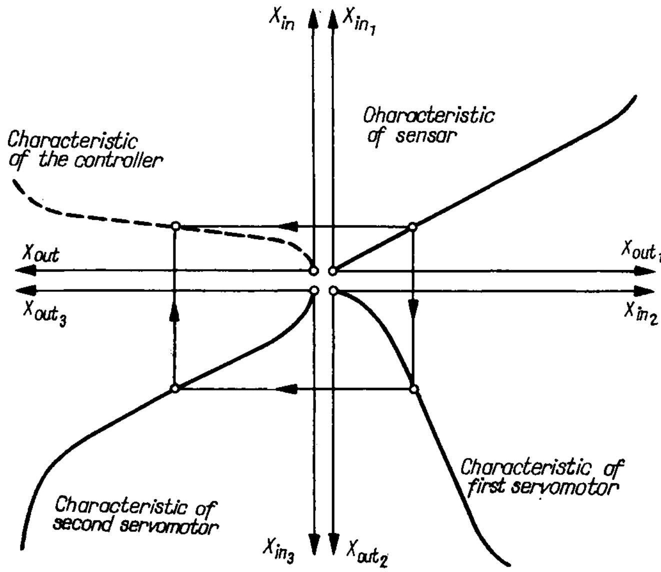


FIG. 73

Let X_{10} be the controlled co-ordinate. Then the segment C (Fig. 70*b*) defines the *static error* or *non-uniformity* of the control for a drop in load from B to A.

If the controller were astatic, then its static characteristic would be a vertical straight line and the static control characteristic would be a horizontal straight line (Fig. 71). If the static error were equal to zero, then the control would be *astatic*.

If we take into account the presence of the setting device of the controller in this construction, then instead of one controller characteristic we would have to consider a family of characteristics and to the control system as a whole there would correspond a family of static characteristics (each curve of this family being related to its tuning value).

The indirect control static characteristic is constructed in exactly the same way, but in this case the control system consists of more than two elements and we first construct the static controller characteristic from the static characteristics of its elements.

The construction can be made directly on the graph if we remember that the output coordinate of an element is the input coordinate of the element which follows it.

EXAMPLE. Let us construct the static characteristic of a controller consisting of a sensor and two servomotors, each of which has feedback. The static characteristics of each of these elements are separately in Fig. 72.

Points of the static controller characteristic as a whole can easily be obtained by arranging the characteristics of the separate elements in the way shown in Fig. 73. The controller characteristic is constructed in the second quadrant.

If there is an astatic element among the controller elements, then the controller characteristic will be a vertical line, and the control system characteristic will be a horizontal line.

A similar construction is easily made also for systems containing feedback. Here it is first necessary to find the static characteristic of each section of the network containing feedback by an exactly similar construction.

When the static controller characteristic has been constructed, from it and the static characteristic of the controlled object we can construct the static control system characteristic in the same way as for the case of direct control described above (Fig. 70). Further, from the static characteristics of the separate elements we can find the steady values of all the generalized co-ordinates for any state, for example, for any fixed value of the load on the controlled object.

As a result of the construction we have described we can find the curves which define the equilibrium values of all the co-ordinates of the system in each possible state (for example, for various values of the load on the object, for various values of the tuning parameters, and so on).

3. The Equations of the System Elements

It follows from the static characteristic constructed in the previous section that when there is a change in the conditions (in the load on the object, the position of the tuning element, etc.) the value of the

coefficient of amplification of the element will change. It will later be shown that other properties of the element, too, depend on these conditions. For this reason we investigate the control process for each possible state separately.*

In such an investigation let us agree to take the equilibrium values of all the generalized coordinates, determined by the method in the previous section, as the zero. If the system was in some equilibrium position and then, as a result of a change in the load on the controlled object, or as a result of action on the controller (change in the tuning), it moved to a new position of equilibrium, then as the zero of all the generalized coordinates we agree to take their values in the position which was the equilibrium position of the system up to the time of the disturbing action.

We denote the values of the coordinates calculated with respect to this zero by ΔX, keeping the same suffix, so that

$$\Delta X_{\text{in}} = X_{\text{in}} - X_{\text{in}\,0}, \quad \Delta X_{\text{out}} = X_{\text{out}} - X_{\text{out}\,0},$$

where X_{in}, X_{out} are the input and output coordinates calculated from the arbitrarily chosen zero, and $X_{\text{in}\,0}$, $X_{\text{out}\,0}$ are the values they have in the investigated state. It is often convenient to change to dimensionless coordinates, which will be denoted by small letters with the same suffix so that

$$x_{\text{in}} = \frac{\Delta X_{\text{in}}}{X^*_{\text{in}}}, \qquad x_{\text{out}} = \frac{\Delta X_{\text{out}}}{X^*_{\text{out}}}$$

where X^*_{in} and X^*_{out} are arbitrarily selected values of the input and output coordinates (basic values*).

The *equation of motion of an element* is the (usually differential) equation which determines the change (in time) of the output coordinate of the element for a given change (in time) in its input coordinate.

* We note at once that it is not possible, from an investigation of the system separately for each possible set of conditions, to draw any conclusions about its properties for big changes in the load. For example, from the fact that, for every value of the load N within the limits $N_1 < N < N_2$, the processes caused by a small change in the load satisfy some technical conditions, it is not possible to deduce that the same technical conditions will be satisfied by the process caused by a large change in the load, even if it does not go outside the limits $N_1 < N < N_2$.

* The basic values are sometimes chosen so as to reduce the number of coefficients which enter into the equation of the control process and are different from unity.

The field of automatic control is so large that it is not possible to form in advance the equations for all the elements which are encountered in practice. Indeed, there is no need to do this, since each time the initial equations are formed new factors, the results of new research, must be taken into account. In order to describe exactly the process for any element we must be guided by previously acquired knowledge. Thus the derivation of the initial equations cannot be performed once and for all and remains an arduous task which the engineer must solve anew each time, taking the specific character of the particular case he is considering into account.

The initial equation for an element usually relates the rate of change of the deviation in the output coordinate $\frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t}$ with the deviations in the values of the input and output coordinates ΔX_{in}, and ΔX_{out}, and is of the form

$$\frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} = F(\Delta X_{\mathrm{in}}, \Delta X_{\mathrm{out}}). \tag{2.1}$$

In some cases, as we shall show later, the initial equation also contains the second derivative of the output coordinate and is of the form:

$$\frac{\mathrm{d}^2\Delta X_{\mathrm{out}}}{\mathrm{d}t^2} = F\left(\Delta X_{\mathrm{in}}, \Delta X_{\mathrm{out}}, \frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t}\right). \tag{2.2}$$

We shall now demonstrate the reasoning used to derive the initial equation, i.e. to determine the function F in equations (2.1) and (2.2), by doing a number of typical problems.

(1) *The equation when an engine (diesel, carburettor, steam) is the controlled object*

We take the change in angular velocity as the generalized output coordinate. For our input coordinate we take the change in position of the control element (which we shall call a damper for convenience*),

$$\Delta X_{\mathrm{in}} = \Delta a.$$

* For example, the throttle of a carburettor engine, the pump lath of a diesel, etc.

The equation of motion of the engine flywheel is of the form

$$I \frac{\mathrm{d}\Delta\omega}{\mathrm{d}t} = \Delta M \tag{2.3}$$

where I is the combined moment of inertia of all of the moving masses of the engine; ΔM is the difference between the change in the torque

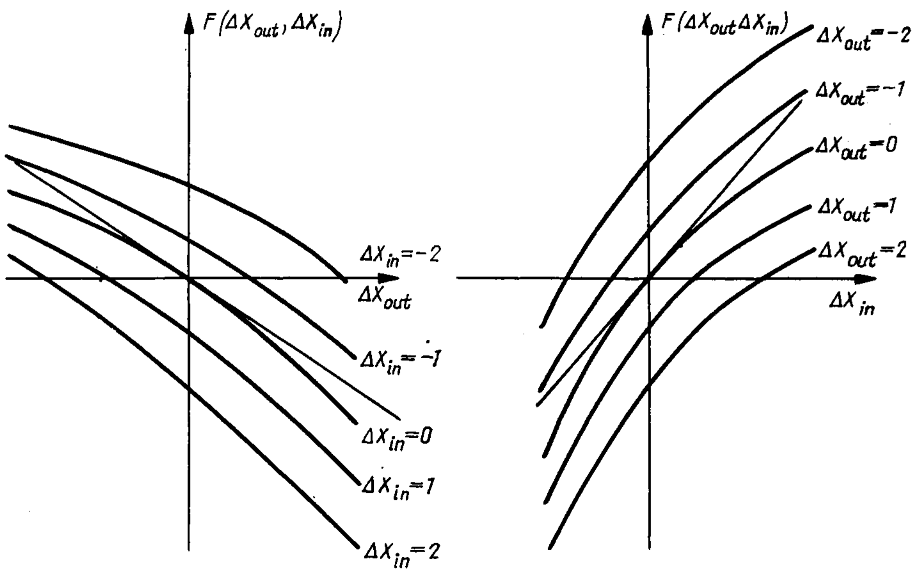


FIG. 74

developed by the engine and the torque caused by resistance. Usually $\Delta M = F(\Delta\omega, \Delta a)$.

We rewrite equation (2.3) in the form

$$I \frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} = F(\Delta X_{\mathrm{out}}, \ \Delta X_{\mathrm{in}}). \tag{2.4}$$

The form of the function $F(\Delta X_{\mathrm{out}}, \Delta X_{\mathrm{in}})$ is determined by the character of the torque characteristics of the engine and by the dependence of external resistances on the angular velocity.

An example of the graph of a function $F(\Delta X_{\mathrm{out}}, \Delta X_{\mathrm{in}})$ is given in Fig. 74.

(2) *The equations of a capacitance (of a unit capacitance element)*

Many elements of automatic control systems can be grouped under the one term capacitance. Although such elements differ externally from each other, the processes in them are subject to the same laws, and are therefore defined by identical equations.

An element is called a capacitance when the following three conditions are fulfilled:

(1) The equilibrium in this element depends on the inflow and outflow of the operating agent being equal;

(2) The element contains only one "reservoir" in which the quantity of operating agent can increase or decrease if the inflow is not equal to the outflow;

(3) The inflow and outflow of the operating agent depend only on the input, or else simultaneously on the input and output coordinates of the element.

A capacitance is said to be *pneumatic* when gas (for example air) is the inflowing and outflowing operating agent, when the reservoir is some volume and when the output coordinate is the pressure in this volume.

Let us consider a volume V and let the gas (air, say) flow into V across a section f_1 and flow out of it across a section f_2. Let the instantaneous value of the rate of flow per second of air across f_1 be equal to Q_1, and the instantaneous value of the rate of flow per second of air across f_2 be equal to Q_2.

Let us agree to attach the suffix 0 to the value of Q_1 and Q_2, and also to f_1 and f_2, at the moment when equilibrium is established, so that

$$Q_{10} = Q_{20}.$$

Where there is no equilibrium, then

$$Q_1 = Q_{10} + \varDelta Q_1 , \quad \text{and} \quad Q_2 = Q_{20} + \varDelta Q_2 ,$$

In time $\mathrm{d}\,t$ the quantity of air contained in V increases or decreases by the amount $\varDelta Q \,\mathrm{d}t = (\varDelta Q_1 - \varDelta Q_2)\,\mathrm{d}\,t$, and therefore the differential of the specific weight γ of the air in the volume V is equal to

$$\mathrm{d}\gamma = \frac{\varDelta Q \,\mathrm{d}t}{V}, \text{ or } V\mathrm{d}\gamma = \varDelta Q \,\mathrm{d}t.$$

Neglecting the change of temperature in the process of small changes in γ and using the equation of state of a gas

$$\frac{p}{\gamma} = RT_{abs} \quad \text{or} \quad \mathrm{d}p = RT_{abs}\,\mathrm{d}\gamma,$$

we obtain

$$\frac{V}{RT_{abs}}\,\mathrm{d}p = \Delta Q \mathrm{d}t,$$

where p is the air pressure in V, R the characteristic constant of air, and T_{abs} is its absolute temperature. Then

$$D\frac{\mathrm{d}p}{\mathrm{d}t} = \Delta Q,$$

where

$$D = \frac{V}{RT_{abs}}$$

Let the air pressure in V at the moment of attaining equilibrium be equal to p_0, so that $p = p_0 + \Delta p$.

Let us denote the deviation in the pressure, the output coordinate, by ΔX_{out}. Then $D\dfrac{\mathrm{d}\Delta X_{out}}{\mathrm{d}t} = \Delta Q$

The value of the difference

$$\Delta Q = \Delta Q_1 - \Delta Q_2 \tag{2.5}$$

depends, generally, speaking, on the deviation in the output and input coordinates. The input co-ordinate can be (Fig. 75) : (1) the section f_1 (or f_2), if this section is altered by means of a valve, damper, etc. ; (2) the pressure p_1 in the space from which the inflow into the volume V comes, if this pressure is the output coordinate of the preceding element ; (3) the pressure p_2 in the space where the outflow from V goes, etc. In all cases,

$$\Delta Q = F(\Delta X_{in}, \ \Delta X_{out})$$

and therefore the equation of motion takes the form

$$\frac{\mathrm{d}\Delta X_{out}}{\mathrm{d}t} = \frac{1}{D}F(\Delta X_{in}, \ \Delta X_{out}). \tag{2.6}$$

This equation is the same as the equation for an engine (2.4) if we replace I by D.

A capacitance is said to be *hydraulic* if the inflowing and outflowing operating agent is a liquid, the reservoir is any volume and the output coordinate is the level of the liquid in this volume.

In the most general case (Fig. 76) when the inflow and outflow of liquid depend on the level in the capacitance, the formula for the inflow per second of liquid across a section f_1 is written in the form

$$Q_1 = s_1 \gamma \sqrt{\frac{2g}{\gamma}} \cdot \sqrt{(p_1 - \gamma h)},$$

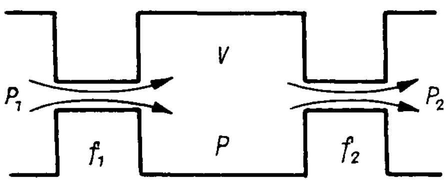


FIG. 75

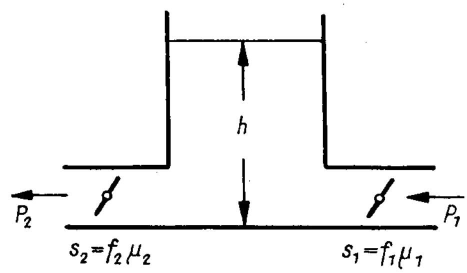


FIG. 76

and that for the outflow per second across a section f_2 in the form

$$Q_2 = s_2 \gamma \sqrt{\frac{2g}{\gamma}} \cdot \sqrt{(\gamma h - p_2)},$$

where $s_1 = f_1 \mu_1$, $s_2 = f_2 \mu_2$; f_1 and f_2 are the flow areas, μ_1 and μ_2 are the flow coefficients and h is the level.

The quantity of liquid in the capacitance changes in time $\mathrm{d}t$ by a quantity

$$\Delta Q \, \mathrm{d}t = (Q_1 - Q_2) \, \mathrm{d}t$$

and therefore

$$\Delta Q \, \mathrm{d}t = \gamma \, \mathrm{d}V,$$

where V is the volume of liquid in the capacitance.

Let V be a function of h. Then

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}h} \frac{\mathrm{d}h}{\mathrm{d}t} = \frac{\Delta Q}{\gamma}.$$

In the simplest case of a cylindrical vessel

$$V = sh, \frac{dV}{dh} = s \quad \text{and} \quad \frac{dh}{dt} = \frac{\Delta Q}{\gamma s},$$

where s is the area of the vessel.

Let $h = h_0 + \Delta X_{out}$ and the input coordinate be the reduced flow area of the inlet choke, $s_1 = s_0 + \Delta X_{in}$. Then

$$\frac{d\Delta X_{out}}{dt} = \frac{\Delta Q}{\gamma s} = \frac{F(\Delta X_{in}, \Delta X_{out})}{\gamma s},$$

or

$$\frac{d\Delta X_{out}}{dt} = \frac{1}{D} F(\Delta X_{in}, \Delta X_{out}),$$

where

$$D = \gamma s$$

and

$$F(\Delta X_{in}, \Delta X_{out}) = \Delta Q =$$

$$= (s_0 + \Delta X_{in}) \gamma \sqrt{\frac{2g}{\gamma}} \sqrt{p_1 - \gamma h_0 - \gamma \Delta X_{out}} \tag{2.7}$$

$$- s_2 \gamma \sqrt{\frac{2g}{\gamma}} \sqrt{\gamma h_0 + \gamma \Delta X_{out} - p_2}.$$

We now consider *thermal* capacitance.

Any body, solid, liquid or gaseous, is said to be a thermal capacitance if heat is supplied to it and taken from it, and the output coordinate is the mean temperature of the body

We completely exclude from our consideration phenomena connected with the distribution of heat waves, and here restrict ourselves to discrete idealized processes only.

Let Θ be the temperature of a body, of mass M and thermal capacity per unit mass C. Let Q_1 and Q_2 be the quantities of heat conducted respectively to and from the body per unit of time, so that

$$Q_1 = Q_{10} + \Delta Q_1$$

and

$$Q_2 = Q_{20} + \Delta Q_2,$$

where $Q_{10} = Q_{20}$ is the quantity of heat conducted to or taken from it during heat equilibrium, then

$$\Delta Q_1 - \Delta Q_2 = \Delta Q$$

is the heat expended per unit of time in the change of temperature of the body.

If $d\theta$ is the change in temperature during a time dt, then

$$MC\, d\theta = \Delta Q\, dt$$

and therefore

$$MC \frac{d\theta}{dt} = \Delta Q,$$

or

$$\frac{d\theta}{dt} = \frac{1}{MC} \Delta Q = \frac{1}{MC} F(\Delta X_{\text{in}}, \Delta X_{\text{out}}). \tag{2.8}$$

Both the heat conducted to the body, ΔQ_1, and the heat taken from it, ΔQ_2, depend on the conditions of heat exchange and, in particular, on the change in the position of the control element of ΔX_{in} which influences the conditions of heat exchange ; in addition, either both quantities ΔQ_1 and ΔQ_2 or at least one of these quantities depend on the temperature change which appears as the output co-ordinate, i.e.

$$\Delta Q = F(\Delta X_{\text{in}}, \Delta X_{\text{out}}),$$

where

$$F(0, 0) = 0.$$

In each actual case the function F is calculated according to the laws of heat transfer and of heat radiation. Comparing (2.6), (2.7) and (2.8) we notice that the equations of pneumatic, hydraulic and thermal capacitance are identical.

The engine which we discussed above can also be looked upon as a capacitance. The equilibrium value of the output coordinate (angular velocity) is determined by the equality of the engine torque and the torque due to resistance. The "reservoir", which, in this case, accumulates the energy used to destroy this equality, is the inertia of the fly wheel. The "inflow of energy" depends on the input and output coordinates, and the "outflow" on the output coordinate only.

(3) *A mechanical centrifugal sensor*

Let us select a point of the sensor, and reduce its mass and all the forces acting on it to this point.

In the general case all the reduced forces may be divided into :

(1) those forces which are functions of the velocity of the output coordinate and which act opposing it in direction ;

(2) those forces which depend only on the output coordinate ;

(3) those forces of external action which include, in particular, forces depending on the input coordinate.

Let X_{out} (the output coordinate) be the deflection of the sensor clutch from the equilibrium position, and X_{in} (the input coordinate) be the deflection of the controlled quantity, measured by the sensor, from its equilibrium value.

In the general case M_{red}, the reduced mass, is a function of the generalized coordinate X_{out}. Let us denote this by $M_{red} = M(X_{out})$. Then the inertial force is equal to

$$M(X_{out}) \frac{d^2 X_{out}}{dt^2} .$$

To the first group of forces belong forces of dry and viscous friction $\bar{f}_1 \left(\frac{dX_{out}}{dt} \right)$ and $f_1 \left(\frac{dX_{out}}{dt} \right)$ respectively.

The forces in the second and third groups can be very diverse linear or non-linear functions of the generalized co-ordinates X_{out} and X_{in}. Let us denote them by $f_2 (X_{in}\ X_{out})$.

We write the differential equation of the sensor, in accordance with D'Alembert's principle, in the form

$$M(X_{out}) \frac{d^2 X_{out}}{dt^2} + \bar{f}_1 \left(\frac{dX_{out}}{dt} \right) + f_1 \left(\frac{dX_{out}}{dt} \right) + f_2 (X_{out}, X_{in}) = 0. \tag{2.9}$$

(4) *The equation of motion of a hydraulic servomotor*

The equation of motion of the piston of a servomotor is determined by the basic law of the non-discontinuity of a jet.

Let us at first neglect the mass of the piston and the external load supported by the piston in action. Then the piston is similar to a thin

film of liquid (Fig. 77*a*). In this case a small displacement of the film will be determined by the equation

$$s\mathrm{d}X_{\text{out}} = fv\mathrm{d}t = \mu f \sqrt{\frac{2g}{\gamma}}\, p_1\, \mathrm{d}t,$$

where s is the area of the surface of the film ; f is the section of the channel along which the oil is brought (this section is altered by the position of the damping element (choke, slide, etc.) ; v is the velocity of the liquid at the section f ; p_1 is the excess pressure ; X_{out} is the change in the level of the liquid (the displacement of the given film) ;

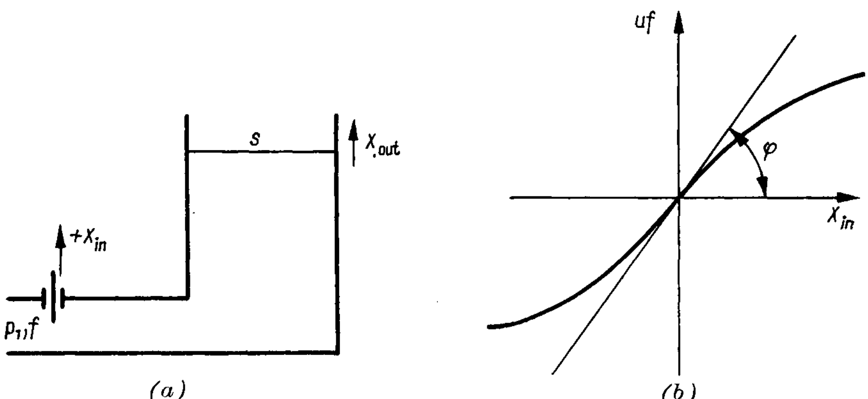


FIG. 77

γ is the specific gravity ; g is the acceleration due to gravity ; μ is the flow coefficient.

If $p_1 = $ const., then this equation can be written :

$$\frac{\mathrm{d}X_{\text{out}}}{\mathrm{d}t} = \text{const}\, \mu f.$$

The effective section of the channel, μf, is a function (Fig. 77*b*) of the increment in the opening of the damping element, X_{in}, i.e. $\mu f = \varphi\,(X_{\text{in}})$, and the equation of motion of the piston takes the form

$$\frac{\mathrm{d}X_{\text{out}}}{\mathrm{d}t} = A\varphi(X_{\text{in}}), \tag{2.10}$$

where

$$A = \frac{1}{s}\sqrt{\frac{2g}{\gamma}}\, p_1.$$

Let us now consider the servomotor which is schematically represented in Fig. 78. We shall suppose that the whole cylinder, both above and below the piston, is filled with oil, and that the slide is so constructed that changes in the cross section in the inflow and outflow ports are always equal in magnitude but opposite in sign.

Now the flow of oil is acted upon by the pressure drop $p_1 - p_x$ or $p_x - p_0$ where p_x is the pressure in the cylinder. The equation of non-discontinuity of a jet leads to the equation

$$p_1 - p_x = p_x - p_0,$$

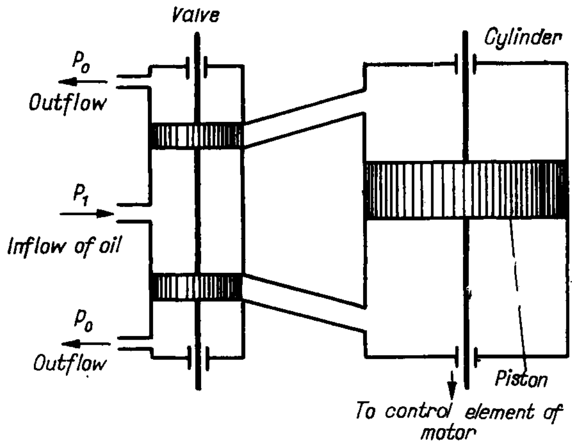


FIG. 78

or

$$p_x = \frac{p_1 + p_0}{2}.$$

Thus the drop which acts upon the flow is equal to

$$p_1 - p_x = p_1 - \frac{p_1 + p_0}{2} = \frac{p_1 - p_0}{2}.$$

Thus, equation (2.10) can also be used as the equation for an ideal two-sided servomotor, if in the formula for A we substitute $\frac{p_1 - p_0}{2}$ for p_1, i.e., if we put in this case

$$A = \frac{1}{s}\sqrt{\frac{g}{\gamma}(p_1 - p_0)}.$$

Figure 79 shows an ordinary flow function $\mu f = \varphi(X_{in})$ for the valve shown in Fig. 78, where X_{in} is the displacement of the valve from its mean position. This curve is symmetrical with respect to the point O. The segment $-O_1 O_1$ corresponds to the "dead space" of the valve (in Fig. 79 this is greatly magnified for the sake of clarity).

We return now to the servomotor with a jet represented schematically in Fig. 80. In this servomotor we cannot change the cross section across which the oil enters the cylinder, but we can change the pressure drop p_1, equal to the difference between the pressure at the output and that at the receiving jet.

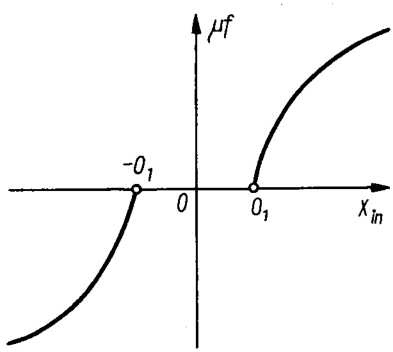


FIG. 79

If we assume that this drop depends only on the position of the jet ΔX_{in}, then the graph of the function $p_1 = f(\Delta X_{in})$ is similar to Fig. 79 with a very small segment $-O_1 O_1$ and with a very steep rise in the characteristic outside this segment.

As before the motion of the piston of the two-sided action servomotor is governed by the equation

$$\mu f \sqrt{\frac{g}{\gamma} p_1}\, \mathrm{d}t = s\mathrm{d}X_{out},$$

but now $\mu f = \text{const}$, and $p_1 = \varphi(X_{in})$.

Thus the equation of motion becomes

$$\frac{\mathrm{d}X_{out}}{\mathrm{d}t} = R\sqrt{\varphi(X_{in})}, \tag{2.11}$$

where

$$R = \frac{\mu f}{s} \sqrt{\frac{g}{\gamma}},$$

s is the area of the piston and f is the cross section at the base of the jet.

When forming the equation of a servomotor we may consider that all those factors which cause a different kind of deflection from that of the equations (2.10) or (2.11) are secondary. Such factors

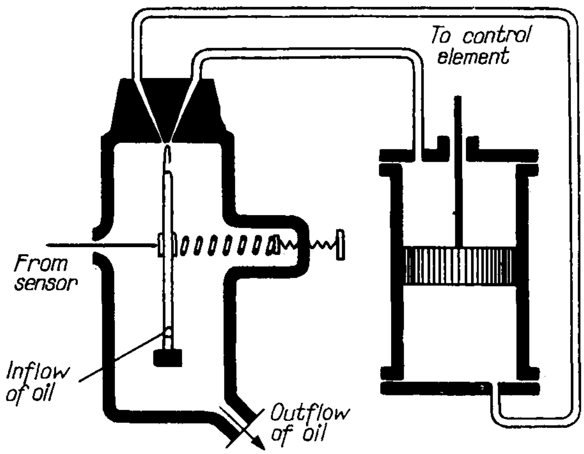


Fig. 80

include the influence, for instance, of the inertia of the piston and of other masses moving with it, of the load supported by the piston during its movement, of the hydraulic resistances on the inlet and outflow lines and so on.

The role of all these factors is that they lessen the pressure drop under the action of which the oil flows through the inflow port.

Let us turn again to the servomotor shown in Fig. 78. In the derivation of its equation let us now take into account the effect of the secondary factors mentioned above.

We denote by $M(X_{\text{out}})$ the mass of all the components, including the control element,* reduced to the piston, by $N(X_{\text{out}})$ the force applied to the stem of the piston, and $\Delta p = \Delta p\left(\frac{\mathrm{d}X_{\text{out}}}{\mathrm{d}t}\right)$ the loss of pressure in the hydraulic resistances in the inflow and outflow lines.

* This reduced mass may, of course, be constant.

The flow of liquid in the cylinder is acted upon by a pressure drop

$$p_1 - p_x - \Delta p - \frac{M(X_{\text{out}}) \dfrac{\mathrm{d}^2 X_{\text{out}}}{\mathrm{d}t^2} + N(X_{\text{out}})}{s}.$$

where

$$\frac{M(X_{\text{out}}) \dfrac{\mathrm{d}^2 X_{\text{out}}}{\mathrm{d}t^2} + N(X_{\text{out}})}{s}$$

is the opposing pressure of the piston.

From the conditions of non-discontinuity of a jet we now have :

$$\mu f \sqrt{\frac{g}{\gamma}} \sqrt{\left[p_2 - p_0 - \Delta p - \frac{1}{s} M(X_{\text{out}}) \frac{d^2 X_{\text{out}}}{dt^2} - \frac{1}{s} N(X_{\text{out}})\right]} = s \frac{\mathrm{d}X_{\text{out}}}{\mathrm{d}t}. \tag{2.12}$$

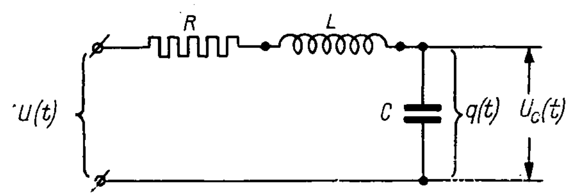


FIG. 81

(5) *An electrical network*

We consider a change in the charge on the plates of a condenser, which is connected in series with a resistance and an inductance (Fig. 81).

Let the change in potential at the terminals (the input coordinate) be given in the form $U(t)$. Then the potential drop across the resistance, inductance and capacitance are given by

$$R \frac{\mathrm{d}q}{\mathrm{d}t}, \quad L \frac{\mathrm{d}^2 q}{\mathrm{d}t^2} \text{ and } \frac{1}{C} q, \text{ respectively,}$$

where $q = CU_C$. The potential drops can, therefore, be expressed in terms of U_C by

$$RC \frac{\mathrm{d}U_C}{\mathrm{d}t}, \quad LC \frac{\mathrm{d}^2 U_C}{\mathrm{d}t^2} \text{ and } U_C \text{ respectively.}$$

Hence,

$$LC\frac{d^2 U_C}{dt^2} + RC\frac{dU_C}{dt} + U_C = U(t).$$

Let

$$U_C = X_{out},\ U(t) = X_{in}.$$

Then the equation of the network will be of the form

$$CL\frac{d^2 X_{out}}{dt^2} + CR\frac{d X_{out}}{dt} + X_{out} = X_{in}. \tag{2.13}$$

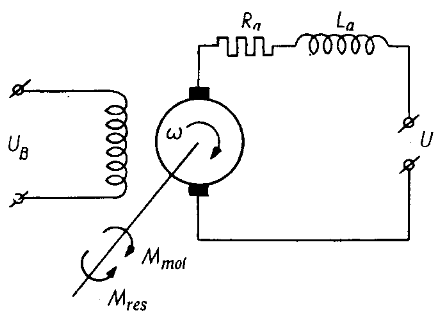


FIG. 82

(6) *A direct current motor*

As an example of the derivation of the equation for an electric motor we consider a direct current motor with independent excitation when the output coordinate is the speed of the armature, and the input coordinate is the change in the potential applied to the armature ; the excitation current being constant (Fig. 82).

The equation of moments is

$$J\frac{d\omega}{dt} = M_{mot} - M_{res},$$

where J is the moment of inertia of the armature, ω is the angular velocity, and M_{mot} and $M_{r.s}$ are the moments of motion and of resistance.

We express the motor moment M_{mot} in terms of the current in the armature coil :

$$M_{mot} = C_{mom}\,\Phi i_a$$

Here C_{mom} is the moment constant of the motor, Φ is the field flux, and i_a is the current in the armature coil.

Let us suppose, in addition, that the moment of the resistance is a linear function of the angular velocity,

$$M_{\text{res}} = a\omega.$$

Then

$$J\frac{d\omega}{dt} = C_{\text{mom}}\Phi i_a - a\omega.$$

The equation of the armature circuit is

$$U = e + L_a\frac{di_a}{dt} + R_a i_a,$$

where U is the potential at the terminals of the motor, e is the e.m.f. induced in the armature ; where L_a and R_a are, respectively, the inductance and resistance of the armature circuit.

The value of e can be expressed as

$$e = C_e\Phi\omega,$$

where C_e is a constant of the motor.

Thus for ω and i_a we obtain the two equations :

$$\left.\begin{aligned} J\frac{d\omega}{dt} &= C_{\text{mom}}\Phi i_a - a\omega, \\ U - C_e\Phi\omega &= L_a\frac{di_a}{dt} + R_a i_a \end{aligned}\right\}$$

Eliminating i_a we find :

$$\begin{aligned} JL_a\frac{d^2\omega}{dt^2} + (R_a J + aL_a)\frac{d\omega}{dt} &+ \\ (aR_a + C_{\text{mom}}C_e\Phi^2)\,\omega &= C_{\text{mom}}\Phi U. \end{aligned} \tag{2.14}$$

(7) *A crossfield electrical amplifier (Fig. 83)*

Neglecting the mutual inductance of the coils 1 and 2, we assume the potential U_B (in coil 1) to be constant and given. Let us form, firstly, the electrical equilibrium equation for coil 2, which receives an input signal U :

$$U = L_2\frac{di_2}{dt} + R_2 i_2,$$

where L_2 and R_2 are the inductance and resistance of the coil, i_2 is the current in it and U is the input potential.

On the linear part of the magnetization characteristic, the potential U_K at the terminals of the short-circuited coil is equal to

$$U_K = k_1' i_1 - k_2' i_2,$$

and for identical coils 1 and 2 ($R_1 = R_2$)

$$U_K = k' (i_1 - i_2),$$

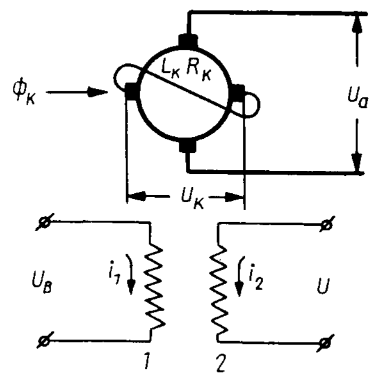


FIG. 83

where $i_1 = \dfrac{U_B}{R_1} = \dfrac{U_B}{R_2}$, the equation for i_2 having been obtained above. The electrical equilibrium equation of the short-circuited circuit is

$$U_K = L_K \frac{di_K}{dt} + R_K i_K,$$

where L_K and R_K are the inductance and resistance of the short circuited circuit and i_K is the current in it.

The output potential U_a is proportional to i_K :

$$U_a = k'' i_K.$$

Eliminating all variables except U_a and U from these equations, and assuming that $\dfrac{di_1}{dt} = 0$ we obtain:

$$L_K L_2 \frac{d^2 U_a}{dt^2} + (L_K R_2 + L_2 R_K) \frac{dU_a}{dt} + R_K R_2 U_a = k'k''(U_B - U).$$

If we let $U_{\mathrm{B}} - U = X_{\mathrm{in}}$ and $U_a = X_{\mathrm{out}}$, we obtain

$$L_{\mathrm{K}} L_2 \frac{\mathrm{d}^2 X_{\mathrm{out}}}{\mathrm{d}t^2} + (L_{\mathrm{K}} R_2 + L_2 R_{\mathrm{K}}) \frac{\mathrm{d}X_{\mathrm{out}}}{\mathrm{d}t} + R_{\mathrm{K}} R_2 X_{\mathrm{out}} = k' k'' X_{\mathrm{in}}. \tag{2.15}$$

4. The Linear Model of an Element. The Linearization of Equations

Among the examples we considered above we encountered elements whose process was defined by a linear differential equation with constant coefficients. This was the case, for example, for electric networks not containing iron (equation (2.13)). The equations of the other elements were non-linear. We can only succeed in obtaining sufficiently general results when such elements are present if we can restrict our consideration to *small disturbances*, i.e. if we can take ΔX_{out} and ΔX_{in} to be small quantities. The words "ΔX_{out} and ΔX_{in} to be small quantities" must here be understood in the following sense : the squares, higher powers and derivatives of these quantities are considerably smaller than their first degree, and are negligible.

Only on these assumptions do we use the general methods of the linear theory of control given in the following chapters. This fact greatly reduces the opportunity for using with cetainty the linear methods of the theory of automatic control in practical calculations. Linear methods cannot validly be used even for small disturbances if the system contains non-linearizable non-linear elements (see below).

Those elements whose non-linear equations can be replaced by linear equations, if only for small disturbances, are said to be *linearizable*. Let us now suppose that the system contains only linear and linearizable elements and that only small disturbances will be considered. The linear equations obtained after supposing that the disturbances are small are called *equations of linear approximation.*

The replacement of the true equation by its linear approximation means, essentially, the substitution of the given element by another element, its *linear model.*

Let us first consider the case when the initial equation is put in the form (2.1). We replace it by the linear equation

$$\frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} = \varphi \Delta X_{\mathrm{in}} - a_1 \Delta X_{\mathrm{out}}. \tag{2.16}$$

Similarly, when the original equation is (2.2), we replace it by the equation

$$\frac{\mathrm{d}^2 \Delta X_{\mathrm{out}}}{\mathrm{d}t^2} = b\Delta X_{\mathrm{in}} - a_1 \Delta X_{\mathrm{out}} - a_2 \frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} . \tag{2.17}$$

The numbers a and b in equation (2.16) and (2.17) are chosen so that the following equalities are satisfied :

$$b = \left[\frac{\partial F}{\partial \Delta X_{\mathrm{in}}} \right]_0 ,$$

$$-a_1 = \left[\frac{\partial F}{\partial \Delta X_{\mathrm{out}}} \right]_0 , \qquad -a_2 = \left[\frac{\partial F}{\partial \dfrac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t}} \right]_0$$

The suffix 0 indicates that the derivative has been taken at the origin of coordinates, i. e. that in the expression for the derivatives the values of $X_{\mathrm{in}\,0}$ and $X_{\mathrm{out}\,0}$ corresponding to the given operating conditions adopted as the zero reading are put :

$$\Delta X_{\mathrm{in}\,0} = \Delta X_{\mathrm{out}\,0} = \left[\frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} \right]_0 = 0.$$

From this definition itself it follows that the construction of a linear model is possible only on condition that the derivatives used have a unique and finite value different from zero. In the contrary case, the element is said to be *non-linearizable.*

Then, in order to find the numerical value of the coefficients a and b it is necessary not only to know the function F, but also to determine by means of a preliminary static solution (see Section 2 of this chapter) the values of X_{in} and X_{out} under the given conditions.

EXAMPLE 1. *The Linear Model of a Capacitance.* We replace the non-linear equation (2.6) of the capacitance by the linear equation

$$\frac{\mathrm{d}\Delta X_{\mathrm{out}}}{\mathrm{d}t} = -a\Delta X_{\mathrm{out}} + b\Delta X_{\mathrm{in}} ,$$

putting

$$-a = \frac{1}{D} \left[\frac{\partial F(\Delta X_{\mathrm{in}}, \Delta X_{\mathrm{out}})}{\partial \Delta X_{\mathrm{out}}} \right]_0 , \quad b = \frac{1}{D} \left[\frac{\partial F(\Delta X_{\mathrm{in}}, \Delta X_{\mathrm{out}})}{\partial \Delta X_{\mathrm{in}}} \right]_0$$

When the function $F(\Delta X_{in}\ \Delta X_{out})$ is given analytically, the derivatives can be calculated directly.

If this function is given by a family of curves, then the values of a and b can be found by constructing the tangent at the point corresponding to the given conditions and determined by the static solution (see Fig. 74).

EXAMPLE 2. *The Linear Model of a Sensor.* We now consider equation (2.9). The function $\bar{f}_1\left(\frac{dX_{out}}{dt}\right)$ expressing the dependence of the forces of dry friction on the velocity has the form shown in Fig. 84.

This function has a discontinuity at the point $\frac{dX_{out}}{dt} = 0$. There is no defined tangent at this point, and for the calculation of dry friction, the sensor is a non-linearizable element.

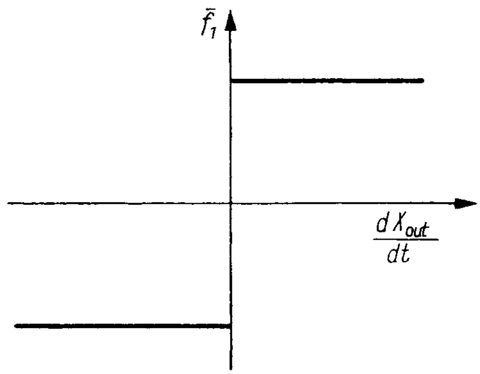


FIG. 84

For this reason the linearization of (2.9) is only possible when the influence of the forces of dry friction can be neglected. If we put $\bar{f}_1 = 0$ in (2.9) and then pass to the increments ΔX_{out} and ΔX_{in} we obtain:

$$M(\Delta X_{out})\frac{d^2\Delta X_{out}}{dt^2} + f_1\left(\frac{d\Delta X_{out}}{dt}\right) + f_2(\Delta X_{out}, \Delta X_{in}) = 0. \tag{2.18}$$

If we put

$$\bar{M} = [M(\Delta X_{out})]_{\Delta X_{out}=0},$$

$$h = \left[\frac{df_1\left(\frac{d\Delta X_{out}}{dt}\right)}{\left(d\,\frac{d\Delta X_{out}}{dt}\right)}\right]_{\frac{dX_{out}}{dt}=0},$$

$$b = -\left[\frac{\partial f_2(\Delta X_{out}\Delta X_{in})}{\partial \Delta X_{in}}\right]_{\substack{\Delta X_{in}=0,\\ \Delta X_{out}=0}},$$

$$c = \left[\frac{\partial f_2(\Delta X_{out}, \Delta X_{in})}{\partial \Delta X_{out}}\right]_{\substack{\Delta X_{in}=0,\\ \Delta X_{out}=0}},$$

where h, b and c are positive numbers, then the equation of the linear model of the sensor takes the form

$$\overline{M}\,\frac{\mathrm{d}^2 \Delta X_{\text{out}}}{\mathrm{d}t^2} + h\,\frac{\mathrm{d}\Delta X_{\text{out}}}{\mathrm{d}t} + c\Delta X_{\text{out}} = \operatorname{sign} b\Delta X_{\text{in}}. \tag{2.19}$$

FIG. 84

We now return to the general case and consider equations (2.16) and (2.17) in detail. They can sometimes be simplified by a transformation of the absolute coordinates to relative (dimensionless) coordinates.

Let ΔX^*_{in} and ΔX^*_{out} be some arbitrarily chosen, but completely defined, values of the coordinates ΔX_{in} and ΔX_{out}, and the ratios

$$x_{\text{in}} = \frac{\Delta X_{\text{in}}}{\Delta X^*_{\text{in}}} \text{ and } x_{\text{out}} = \frac{\Delta X_{\text{out}}}{\Delta X^*_{\text{out}}}$$

be the *relative* or *dimensionless* values of the input and output coordinates.

We note that

$$\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} = \frac{1}{\Delta X^*_{\text{out}}}\,\frac{\mathrm{d}\Delta X_{\text{out}}}{\mathrm{d}t};$$

$$\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} = \frac{1}{\Delta X^*_{\text{out}}}\,\frac{\mathrm{d}^2 \Delta X_{\text{out}}}{\mathrm{d}t^2}$$

and so on. Putting these values of x_{in}, x_{out}; $\mathrm{d}x_{\text{out}}/\mathrm{d}t$, $\mathrm{d}^2 x_{\text{out}}/\mathrm{d}t^2$ in equations (2.16) and (2.17) we obtain the equation of the linear model in dimensionless coordinates.

Equation (2.16) becomes

$$\Delta X^*_{\text{out}}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} = b\Delta X^*_{\text{in}}\,x_{\text{in}} - a_1 \Delta X^*_{\text{out}}\,x_{\text{out}}, \tag{2.20}$$

and equation (2.17) becomes

$$\Delta X^*_{\text{out}}\,\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} = b\Delta X^*_{\text{in}}\,x_{\text{in}} - a_1 \Delta X^*_{\text{out}}\,x_{\text{out}} - a_2 \Delta X^*_{\text{out}}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}. \tag{2.21}$$

Let us first suppose that the coefficient a_1 is different from zero. Then there are two possible forms in which the equations in relative coordinates of the linear model can be put.

The first form of the equation is obtained if we divide equation (2.20) or (2.21) by the coefficient of x_{out}, i. e. by $a_1 \Delta X^*_{\text{out}}$:

$$\frac{1}{a_1}\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = \frac{b}{a_1}\frac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}} x_{\text{in}}, \tag{2.22}$$

$$\frac{1}{a_1}\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + \frac{a_2}{a_1}\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = \frac{b}{a_1}\frac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}} x_{\text{in}}. \tag{2.23}$$

The second form of the equation is obtained* if we divide equation (2.20) or (2.21) by the coefficient of x_{in}, i.e. by $b\Delta X^*_{\text{in}}$:

If the coefficient $a_1 = 0$, then the first form of the equation is obtained in the same way, except that we divide the required equation by the coefficient of $\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}$:

$$\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} = b\frac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}} x_{\text{in}}, \tag{2.26}$$

$$\frac{1}{a_2}\frac{\mathrm{d}^2 X_{\text{out}}}{\mathrm{d}t^2} + \frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} = \frac{b}{a_2}\frac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}} x_{\text{in}}. \tag{2.27}$$

The second form of the equation for the case $a_1 = 0$ does not differ from (2.24) or (2.25) ; it is only necessary to put $a_1 = 0$ in (2.24) or (2.25).

Let us now recall that the numbers ΔX^*_{in} and ΔX^*_{out}, which were introduced in the transformation from absolute to relative co-

* Sometimes a third form of the equation is used, obtained from the equations (2.20) or (2.21) by dividing them by the coefficient of the highest derivative. This form of the equation is less convenient and will not be used in this book.

$$\frac{\Delta X^*_{\text{out}}}{b\Delta X^*_{\text{in}}}\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + \frac{a_1}{b}\frac{\Delta X^*_{\text{out}}}{\Delta X^*_{\text{in}}} x_{\text{out}} = x_{\text{in}}, \tag{2.24}$$

$$\frac{\Delta X^*_{\text{out}}}{b\Delta X^*_{\text{in}}}\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + \frac{a_2}{b}\frac{\Delta X^*_{\text{out}}}{\Delta X^*_{\text{in}}}\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + \frac{a_1}{b}\frac{\Delta X^*_{\text{out}}}{\Delta X^*_{\text{in}}} x_{\text{out}} = x_{\text{in}} \tag{2.25}$$

ordinates, may be chosen arbitrarily. This arbitrariness may be used to simplify the equations in relative coordinates of the linear model in both the first and second form.

Thus, for example, if in equations (2.22) and (2.23) we choose ΔX^*_{in} and ΔX^*_{out} so that $\dfrac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}} = \dfrac{a_1}{b}$, then the coefficients of x_{in} will be equal to unity, and the equations (2.22) and (2.23) will reduce to the form

$$\frac{1}{a_1}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = x_{\text{in}}$$

and

$$\frac{1}{a_1}\,\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + \frac{a_2}{a_1}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = x_{\text{in}},$$

and the equations (2.24) and (2.25) to the form

$$\frac{\Delta X^*_{\text{out}}}{b\Delta X^*_{\text{in}}}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = x_{\text{in}}$$

and

$$\frac{\Delta X^*_{\text{out}}}{b\Delta X^*_{\text{in}}}\,\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + \frac{a_2}{b}\,\frac{\Delta X^*_{\text{out}}}{\Delta X^*_{\text{in}}}\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = x_{\text{in}}.$$

In the formation of the equations of all the elements of the system one and the same coordinate enters into the various equations. Of course having selected some base in order that a coefficient in one of the equations can be taken as unity, we can no longer dispose of this base in the other equations.

Let us consider in more detail equations (2.22) and (2.23), since only the first form of the equations will be used later in this book. In both equations the quantity x_{out} is dimensionless. Hence the other terms in (2.22) and (2.23) must also be dimensionless.

We consider first of all the term $\dfrac{b}{a_1}\,\dfrac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}}\,x_{\text{in}}$, on the right-hand side of these equations. The whole derivative must be dimensionless. The quantity x_{in} is dimensionless, and therefore the quantity $\dfrac{b}{a_1}\,\dfrac{\Delta X^*_{\text{in}}}{\Delta X^*_{\text{out}}}$ also has no dimensions. We denote this dimensionless quantity by the letter k and call it the *coefficient of amplification of the linear model of the element.*

Let us consider any new position of equilibrium of the system, different from that which was taken as the origin of coordinates. Let us assume, for example, that the load and tuning have completely defined values, corresponding to the given conditions, and that the control process is caused by a small change in the tuning or the load. Then for the new position of equilibrium, which corresponds to the new value of the load or tuning, the relative deviations of input and output coordinates of the element are different from zero. Let them be equal to $x_{\text{in}} = x_{\text{in}\,0}$ and $x_{\text{out}} = x_{\text{out}\,0}$. These steady values $x_{\text{in}\,0}$ and $x_{\text{out}\,0}$ are determined by the static solution of the system (see Section 2 of this chapter).

In the position of equilibrium all the derivatives of x_{out} are equal to zero, and equations (2.22) and (2.23) reduce to the form

$$x_{\text{out}\,0} = kx_{\text{in}\,0} \quad \text{or} \quad k = \frac{x_{\text{out}\,0}}{x_{\text{in}\,0}}.$$

*The coefficient of amplification of the linear model of an element is equal to the ratio of the equilibrium value of the output coordinate to the equilibrium value of the input coordinate.**

The left-hand side of (2.22) contains the dimensionless term x_{out}. The term $\dfrac{1}{a_1}\dfrac{dx_{\text{out}}}{dt}$, in (2.22) and the term $\dfrac{a_2}{a_1}\dfrac{dx_{\text{out}}}{dt}$ in (2.23) must also be dimensionless. But $\dfrac{dx_{\text{out}}}{dt}$ has the dimensions $\dfrac{1}{\text{sec}}$ and therefore $\dfrac{1}{a_1}$ and $\dfrac{a_2}{a_1}$ have the dimensions of time.

In (2.22) let us put $\dfrac{1}{a_1} = T$ and in (2.23) $\dfrac{a_2}{a_1} = T_k$.

In (2.23) the term $\dfrac{1}{a_1}\dfrac{d^2 x_{\text{out}}}{dt^2}$ is also dimensionless, but $\dfrac{d^2 x_{\text{out}}}{dt^2}$ has the dimensions $\dfrac{1}{\text{sec}^2}$, and therefore the dimensions of $\dfrac{1}{a_1}$ in this case are sec². In (2.23) we put $\dfrac{1}{a_1} = T'^2$.

* The value of k is determined by the ratio of the dimensionless coordinates. Hence k depends on the choice of the base.

With this notation the equations of the linear model in their first form reduce to

$$\boxed{T\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}+x_{\text{out}}=kx_{\text{in}},} \tag{2.28}$$

$$\boxed{T'^2\frac{\mathrm{d}^2x_{\text{out}}}{\mathrm{d}t^2}+T_k\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}+x_{\text{out}}=kx_{\text{in}}} \tag{2.29}$$

The quantities T' and T, having the dimensions of time, are called the *time constants* of the element. The constant T_k, also having the dimensions of time, is called the *damping time constant.*

The coefficient of amplification k always depends on ΔX^*_{in} and ΔX^*_{out}, but the time constants T and T_k are independent of them.

In all of this discussion it has been assumed that a_1, a_2 and b are positive numbers. In this case all the T and k are also positive.

If any of the coefficients a_1, a_2 or b is negative, then in the formulae for the T and k we introduce the absolute value of this coefficient, so that the T and k remain positive as before, but the corresponding sign in (2.28) and (2.29) is changed.

In an exactly similar way the linear model for other linearizable element is constructed.

5. The Classification of Linear Models of Elements. The Inherent Operator and the Action Operators. Typical Elements (Stages)

By deriving the equations of linear approximation of various elements in the same way as in the previous section, we obtain equations of the form*

$$T\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}\pm x_{\text{out}}=kx_{\text{in}},$$

$$T'^2\frac{\mathrm{d}^2x_{\text{put}}}{\mathrm{d}t}\pm T_k\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t}\pm x_{\text{out}}=kx_{\text{in}},$$

* The linearization of the equations derived in Section 3 as examples leads to the first two types of equation of linear approximation.

$$T\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} \pm x_{\text{out}} = kx_{\text{in}} \pm \varrho\frac{\mathrm{d}x_{\text{in}}}{\mathrm{d}t},$$

$$T'^2\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} \pm T_k\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} \pm x_{\text{out}} = kx_{\text{in}} \pm \varrho\frac{\mathrm{d}x_{\text{in}}}{\mathrm{d}t} \pm s\frac{\mathrm{d}^2 x_{\text{in}}}{\mathrm{d}t^2}.$$

We lay down the initial conditions for the x_{out} to be equal to zero, in the third of these equations we put $x_{\text{in}}(0) = 0$, and in the fourth, in addition, $x_{\text{in}}(0) = 0$. Then the Laplace transform* of the output coordinate of any element, defined by such equations, is connected with the Laplace transforms of the input coordinate by the relation

$$d(p)\,L[x_{\text{out}}] = k(p)\,L[x_{\text{in}}],$$

where $L[x_{\text{out}}]$ and $L[x_{\text{in}}]$ are the Laplace transforms of x_{out} and x_{in}, p is a complex number, and $d(p)$ and $k(p)$ are polynomials in p.

The polynomials $d(p)$ and $k(p)$ may be obtained in the following way: in the equations of linear approximation we replace $\frac{\mathrm{d}x}{\mathrm{d}t}$ by $pL[x]$; $\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}$ by $p^2L[x]$ and so on, and take $L[x]$ outside the brackets. Then the polynomials remaining inside the brackets will be equal to $d(p)$ and $k(p)$.

Thus, for example, with zero initial conditions, applying the Laplace transform to the equation

$$T\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = kx_{\text{in}},$$

we obtain

$$TpL[x_{\text{out}}] + L[x_{\text{out}}] = kL[x_{\text{in}}],$$

or

$$(Tp+1)\,L[x_{\text{out}}] = kL[x_{\text{in}}].$$

In this case

$$d(p) = Tp + 1 \text{ and } k(p) = k.$$

As a second example we consider the equation

$$T'^2\frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + T_k\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = kx_{\text{in}} + \varrho\frac{\mathrm{d}x_{\text{in}}}{\mathrm{d}t}.$$

* See Appendix I for the Laplace transformation.

After taking its Laplace transform, we can write it in the form

$$T'^2 p^2 L[x_{\text{out}}] + T_k pL[x_{\text{out}}] + L[x_{\text{out}}] = kL[x_{\text{in}}] + \varrho pL[x_{\text{in}}],$$

or

$$(T'^2 p^2 + T_k p + 1) L[x_{\text{out}}] = (\varrho p + k) L[x_{\text{in}}].$$

In this case $d(p) = T'^2 p^2 + T_k p + 1$ and $k(p) = \varrho p + k$.

The polynomial $d(p)$ is called *the inherent operator of the element,** and $k(p)$ is called the *action operator* for the element or the *operator coefficient of amplification.*

The elements that are most frequently encountered have the following inherent operators :

$$Tp + 1;\ T'^2 p^2 + T_k p + 1;\ T'^2 p^2 + 1;\ Tp \text{ and } Tp - 1.$$

They have a special significance in control theory, and separate names and an agreed notation are given to them.

An element for which $d(p) = Tp + 1$ will be called *single-capacitance* and will be denoted by a square □.

An element with $d(p) = T'^2 p^2 + T_k p + 1$ will be called *oscillatory* and will be denoted by a rectangle ▭.

An element with $d(p) = T'^2 p^2 + 1$ will be called *conservative* and will be denoted by a shaded rectangle ▨.

An element with $d(p) = Tp$ will be called *astatic* and will be denoted by a circle ○.

Finally, an element with $d(p) = Tp - 1$ will be called *unstable.* We will denote it by a triangle ▽.

The action operators $k(p)$ are most often encountered in the following forms :

$$k,\ k + \varrho p \text{ and } k + \varrho p + sp^2.$$

* The term "operator" is used here as a result of the fact that we may introduce the polynomials $d(p)$ and $k(p)$ without using a Laplace transform, but by the introduction of the operator form of writing differential equations. If we put $\frac{dx}{dt} = px$, $\frac{d^2 x}{dt^2} = p^2 x$, then, for example, the equation

$$T'^2 \frac{d^2 x_{\text{out}}}{dt^2} + T_k \frac{dx_{\text{out}}}{dt} + x_{\text{out}} = kx_{\text{in}} + \varrho \frac{dx_{\text{in}}}{dt}$$

may be written

$$T'^2 p^2 x_{\text{out}} + T_k px_{\text{out}} + x_{\text{out}} = kx_{\text{in}} + \varrho px_{\text{in}},$$

or

$$(T'^2 p^2 + T_k p + 1) x_{\text{out}} = (k + \varrho p) x_{\text{in}}, \quad d(p) x_{\text{out}} = k(p) x_{\text{in}}.$$

Operators with negative signs before the ϱ and s are also found, but much less often.

In the case $k(p) = k =$ const, the actions on the element are called *static*. If $k(p) = k + \varrho\, p$, then, in addition to static action, there is also *first derivative action* and in the case $k(p) = k + \varrho\, p + sp^2$, both *first and second derivative action*.

Elements for which $d\,(p)$ and $k(p)$ are of the above forms are called typical elements or stages*.

6. The Transfer Function of the Linear Model of a System. Its Formation from the Equations of the Linear Models of its Elements

(a) *The concepts of a transfer function*

From the equations of all the stages it is always possible to elim inate all the input coordinates, since the input coordinate of each element can be expressed in terms of the output coordinates of the other elements of the system. As a result, the system of equations of the control process will consist of n equations, connecting n generalized co-ordinates (here n is the number of degrees of freedom which are taken into account). The automatic control process is described by the aggregate of these equations.

In the most general case several input coordinates (the element can have several inputs) may act on any of the element. Moreover, all or some of the coefficients k may be polynomials in p, and in the most general case the equations of motion of the linear model of the system, after Laplace transformation* take the form :

$$\left.\begin{array}{l} d_1\,(p)\,L\,[x_1] + k_{12}\,(p)\,L\,[x_2] + \ldots + k_{1n}\,(p)\,L\,[x_n] = L\,[f_1\,(t)], \\ k_{21}\,(p)\,L\,[x_1] + d_2\,(p)\,L\,[x_2] + \ldots + k_{2n}\,(p)\,L\,[x_n] = L\,[f_2\,(t)], \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ k_{n1}\,(p)\,L\,[x_1] + k_{n2}\,(p)\,L\,[x_2] + \ldots + d_n\,(p)\,L\,[x_n] = L\,[f_n\,(t)]. \end{array}\right\} \quad (2.30)$$

* In the literature on the subject other names for typical stages are used. Thus, for example, a single-capacitance stage is sometimes called aperiodic, a static neutral stage, a stage with the inherent operator $T'^2\,p^2 + T_k\,p + 1$ is sometimes called oscillatory only if the inequality $2T' > T_k$ is satisfied, and so on. In this book only the italicized terms above will be used.

* See Appendix I (pp. 495—6).

In every particular case the inherent operators $d(p)$ and the action operators $k(p)$ are different, and some of the $k(p)$ can be identically equal to zero.

In this system each equation corresponds to one of the system element. Let us confine ourselves for the present to the case when only the external action $f_1(t)$ is different from zero, that is, $f(t)$ acts only on one element of the system, to which in this case we attach the suffix 1.

In order to find the Laplace transform of any of the generalized coordinates x_j, it is necessary to solve the system of algebraic equations (2.30) for $L[x_j]$:

$$L(x_j] = \frac{\Delta_j(p)}{\Delta(p)} L[f(t)].$$

In this equation $\Delta(p)$ is the determinant of the system.

$$\Delta(p) = \begin{vmatrix} d_1(p) & k_{12}(p) \ldots k_{1n}(p) \\ k_{21}(p) & d_2(p) \ldots k_{2n}(p) \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ k_{n1}(p) & k_{n2}(p) \ldots d_n(p) \end{vmatrix}, \tag{2.31}$$

and $\Delta_j(p)$ is the algebraic complement of the element lying in the first row and the jth column.

The equation

$$\Delta(p) = 0 \tag{2.32}$$

is called the *characteristic equation.*

The distribution of its roots solves the question of the stability of the system*.

The function

$$\Phi(p) = \frac{\Delta_j(p)}{\Delta(p)}, \tag{2.33}$$

equal to the ratio $\frac{L[x_j]}{L[f(t)]}$ is called the *transfer function of the closed system* for the coordinate x_j and for the given action upon the first element.

* See Chapter III.

With a change of the coordinate we are considering, or of the point of application of the disturbance the form of the transfer function changes, but only because of a change in the numerator. The denominator, and also of course the characteristic equation of the system, is not changed.

When the process is analysed from the equations of the elements, the transfer function of the system lies at the root of the analysis.

The transfer function of a system can be formed immediately from the equations of its elements, without the formation and row expansion of the above determinants. In order to do this we must introduce the preliminary concept of the transfer function of an element and of more complex parts of the system.

(b) *The transfer function of an element*

From the Laplace-transform equation of the element

$$d(p)\, L[x_{\text{out}}] = k(p)\, L[x_{\text{in}}] \tag{2.34}$$

it follows that

$$\frac{L[x_{\text{out}}]}{L[x_{\text{in}}]} = \frac{k(p)}{d(p)} . \tag{2.35}$$

This ratio, by analogy with the system as a whole, is called the *transfer function of the system element*, and we denote it by $W(p)$.

(c) *The transfer function of a sequential open circuit of elements*

We consider an open circuit of n elements, sequentially acting on one another, so that the input coordinate of each element apart from the first is the output coordinate of the previous element (Fig. 85).

Writing out the transfer functions for all the elements of this system we obtain

$$\frac{L[x_1]}{L[x_{\text{in}}]} = \frac{k_1(p)}{d_1(p)} = w_1(p); \ldots ; \frac{L[x_{\text{out}}]}{L[x_{n-1}]} = \frac{k_n(p)}{d_n(p)} = w_n(p).$$

Eliminating in succession all the $L[x]$, except $L[x_{\text{in}}]$ and $L[x_{\text{out}}]$, we obtain the relation between $L[x_{\text{in}}]$ and $L[x_{\text{out}}]$, i.e. the transfer

function for this elementary circuit :

$$\frac{L[x_{\text{out}}]}{L[x_{\text{in}}]} = W(p) = w_1(p)\, w_2(p) \ldots w_n(p), \tag{2.36}$$

or

$$W(p) = \prod_{j=1}^{j=n} \frac{k_j(p)}{d_j(p)} = \frac{K(p)}{D(p)}\,, \tag{2.37}$$

where*

$$K(p) = \prod_{j=1}^{j\geq n'} k_j(p), \text{ and } D(p) = \prod_{j=1}^{j=n} d_j(p),$$

or

$$\boxed{W(p) = \prod_{j=1}^{j=n} w_j(p).} \tag{2.38}$$

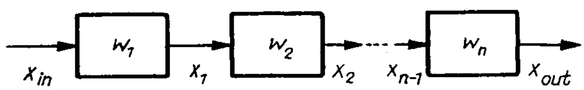


FIG. 85

As distinct from a closed system, the transfer function of an open loop system is denoted by $W(p)$.

Thus, *the transfer function of an open circuit of n sequentially connected elements is equal to the product of the transfer functions of these elements.*

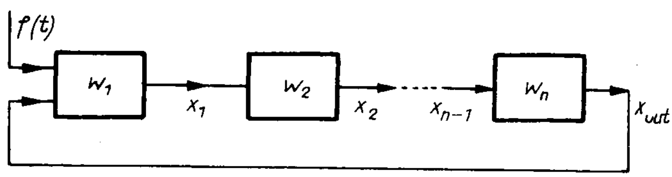


FIG. 86

(d) *The transfer function of a closed single-loop system*

We now consider a closed system in which only the output co-ordinate of the previous element acts at the input of each element. Such a system is called a *single-loop system* (Fig. 86).

* The notation $\prod$ denotes the product of all the quantities standing after it. Sometimes in considering a product of n factors instead of $\prod_{j=1}^{j=n}$ we shall write simply $\prod$.

Let us suppose that the external action is applied to the first element and that it is required to find the Laplace transform for the coordinate x_k.

We open the loop at the input of the first element (Fig. 87*a*). Then from our assumptions, the separate networks of elements from the first to the k-th and from the $(k + 1)$-th to the n-th can be replaced by equivalent elements, whose transfer functions $W_1(p)$ and $W_2(p)$ we calculate from the formula derived above (2.38) :

$$W_1(p) = \prod_{j=1}^{j=k} w_j(p); \quad W_2(p) = \prod_{j=k+1}^{j=n} w_j(p).$$

The circuit for this system is shown in Fig. 87*b*.

We now return to the closed system (Fig. 86).

To the relations

$$W_1(p) = \frac{L[x_k]}{L[x_{\text{in}}]} \tag{2.39}$$

$$W_2(p) = \frac{L[x_n]}{L[x_k]} \tag{2.40}$$

resulting from the definition of the transfer function, we add the conditions of the closed system*

$$x_{\text{in}} = -x_n + f(t)$$

or

$$L[x_{\text{in}}] = -L[x_n] + L[f(t)]. \tag{2.41}$$

Eliminating the two variables $L[x_{\text{in}}]$ and $L[x_n]$ from (2.39), (2.40) and (2.41) we obtain :

$$L[x_k] = \frac{W_1(p)}{1 + W_1(p)W_2(p)} L[f(t)].$$

The required transfer function of the single-loop system is therefore equal to

$$\boxed{\Phi(p) = \frac{L[x_k]}{L[f(t)]} = \frac{W_2(p)}{1 + W_1(p)W_2(p)}} \tag{2.43}$$

* The minus sign in front of x_n indicates that when the loop is closed the control element must cause an increase in the controlled quantity if it is below the desired value at that instant, i. e. the action must change sign at one point (or at an odd number of points) somewhere in the circuit.

The numerator of the transfer function of a single-loop system is equal to the product of the transfer functions of the elements lying between the point of application of the disturbance and the co-ordinate being considered, and the denominator is equal to the product, increased by one, of the transfer functions of all elements in the system.

The characteristic equation of this system

$$1 + W_1(p)\, W_2(p) = 0$$

can be put in the form

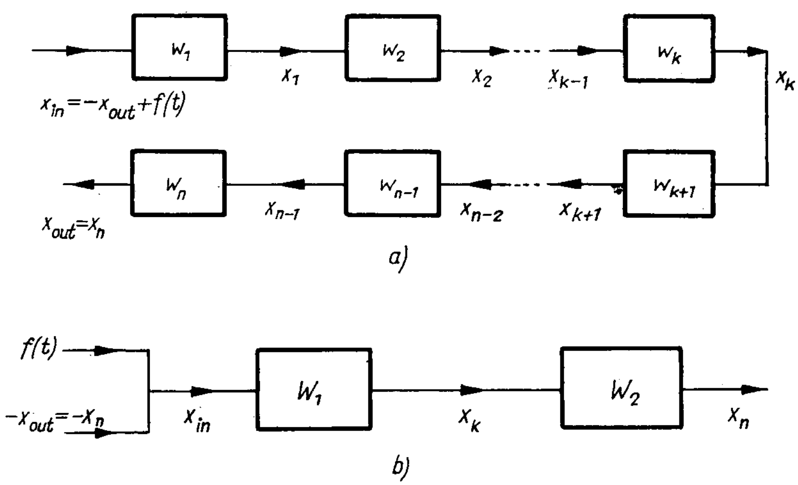


FIG. 87

$$\frac{\Pi d_j(p) + \Pi k_j(p)}{\Pi d_j(p)} = 0$$

or

$$\boxed{D(p) + K(p) = 0.} \tag{2.43}$$

Thus, *the left-hand side of the characteristic equation of a closed single-loop system is equal to the sum of two terms. The first term is the product of all the inherent operators of the elements, and the second is the product of the operator coefficients of amplification of all the elements.*

(e) *The transfer function of a system containing internal feedback*

We confine ourselves to cases when the feedback does not include the element to which the disturbance is applied or an element whose output coordinate is unknown (Fig. 88*a*). Let us suppose that the

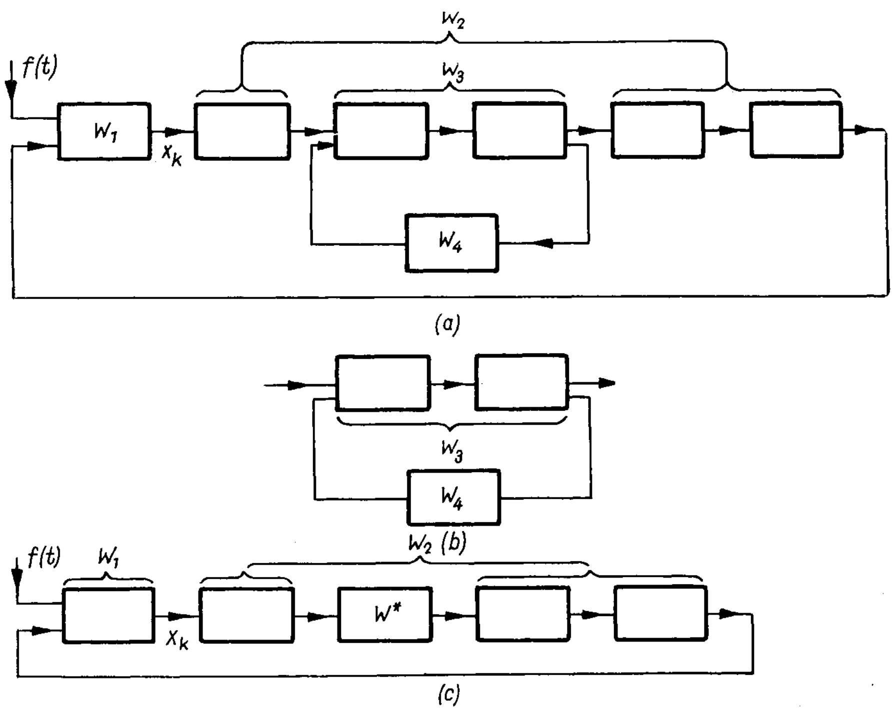


FIG. 88

feedback transfer function and the transfer functions of all the elements are given. It is required to determine the transfer function of the system as a whole. To this end, we first "exclude" from the system the feedback and the stages shunted by it (Fig. 88*b*). The transfer function of the "excluded" part of the system can be determined from formula (2.42), since the excluded part of the system is itself a single-loop system. If $W_3(p)$ is the product of the transfer functions of the elements shunted by feedback, then the transfer function of

the "excluded part" of the system, $W^*(p)$, is equal to

$$W^*(p) = \frac{W_3(p)}{1 + W_3(p)\, W_4(p)},$$

where $W_4(p)$ is the transfer function of the feedback element itself.

In view of this, the total "excluded part" of the circuit can be replaced by an equivalent element having the same transfer function.

As a result the initial non-single-loop system (Fig. 88*a*) is replaced by an equivalent single-loop system (Fig. 88*b*) having a transfer function which may be found from the formula (2.42) :

$$\Phi(p) = \frac{W_1(p)}{1 + W_1(p)\, W_2(p)\, W^*(p)}$$

or

$$\Phi(p) = \frac{W_1(p)}{1 + \dfrac{W_1(p)\, W_2(p)\, W_3(p)}{1 + W_3(p)\, W_4(p)}} =$$

$$= \frac{W_1(p)\,[1 + W_3(p)\, W_4(p)]}{1 + W_3(p)\, W_4(p) + W_1(p)\, W_2(p)\, W_3(p)}.$$

In these formulae, $W^*(p)$ does not enter in W_1.

In more complicated cases also the transfer function and characteristic equation of the system are determined by similar computations using the transfer functions of the elements of the system or their equations.

7. The Statics of the Linear Model of an Automatic Control System. The Transfer Functions of Static and Astatic Systems

The concept of a "transfer function" is closely connected with the question of the determination of some static properties of the linear model of an automatic control system.

In Section 6 it was shown that the Laplace transforms $L[x_k]$ and $L[f(t)]$ are connected by the relation

$$L[x_k] = W(p)\, L[f(t)]. \tag{2.44}$$

From the theory of Laplace transforms it is known that*

* See Appendix I (p. 483).

$$\lim_{t\to\infty} x(t) = \lim_{p\to 0} pL[x(t)] \,.$$

Let $x_{k_{\lim}}$ be the value of the co-ordinate which is set up as a result of the control process (i.e. $\lim\limits_{t\to\infty} x_k(t)$ and $f_{\lim} = \lim\limits_{t\to\infty} f(t)$). Multiplying both sides of (2.44) by p and passing to the limit as $p \to 0$ we obtain

$$\lim_{t\to\infty} x_k(t) = \lim_{p\to 0} W(p)\, pL[f(t)]$$

or

$$x_{k_{\lim}} = W(0) f_{\lim}, \tag{2.45}$$

where $f_{\lim} = \lim\limits_{t\to\infty} f(t)$ is steady state (static) value of $f(t)$.

Let $x_k = x_1$ be the controlled coordinate, and $f(t)$ the load on the controlled object with a transfer function W_1 from $f(t)$ to x_1. If the controller is not included, then the deviation in x_1 caused by the disturbance $f(t)$ will be equal to

$$x_{1_{\lim}} = W_1(0) f_{\lim} = x^*_{1_{\lim}}.$$

When the controller is included

$$x_{1_{\lim}} = \Phi(0) f_{\lim},$$

where

$$\Phi(0) = \left[\frac{W_1(p)}{1 + W_1(p) W_p(p)}\right]_{p=0}.$$

Here $W_p(p)$ is the transfer function of the controller as a whole.

Therefore

$$x_{1_{\lim}} = \frac{W_1(0)}{1 + W_1(0) W_p(0)} f_{\lim}$$

or

$$\boxed{x_{1_{\lim}} = \frac{x^*_{1_{\lim}}}{1 + W_1(0) W_p(0)}.}$$

The quantity $x_{1\,\mathrm{lim}}$ determines the deviation of the controlled coordinate in the new steady conditions (for the same load with the controller operative) i.e. the *static error*.

Thus, *the inclusion of the controller reduces the static error by* $1 + W_1(0)\,W_p(0)$ *times*.

Let $W_1(0) = \dfrac{K_1(0)}{D_1(0)}$ and $W_p(0) = \dfrac{K_p(0)}{D_p(0)}$. Then

$$x_{1\,\mathrm{lim}} = \frac{K_1(0)\,D_p(0)}{D_1(0)\,D_p(0) + K_1(0)\,K_p(0)}\,f_{\mathrm{lim}} \tag{2.46}$$

or

$$x_{1\,\mathrm{lim}} = \frac{D_1(0)\,D_p(0)}{D_1(0)\,D_p(0) + K_1(0)\,K_p(0)}\,x^*_{1\,\mathrm{lim}}. \tag{2.46'}$$

As is clear from (2.46), the system can be made astatic (i.e. such that $x_{1\,\mathrm{lim}} = 0$ for any $f_{\mathrm{lim}} \neq 0$). In order to do this, it is necessary that $D_p(0) = 0$.

Hence, *any automatic control system is astatic only in the case when the inherent operator of the controller, $D_p(p)$, contains p as a factor.*

From (2.46′) it follows that *the greater the total coefficient of amplification $K_1(0)\,K_p(0)$ of the open system, the smaller is the error compared to that which would occur without a controller.*

Let

$$W_1 = \frac{k_1(p)}{d_1(p)}$$

and

$$\Phi(p) = \frac{\dfrac{k_1(p)}{d_1(p)}}{1 + \dfrac{K(p)}{D(p)}} = \frac{\bar{D}(p)}{D(p) + K(p)}\,k_1(p),$$

where

$$D(p) = \Pi\, d_j(p); \quad K(p) = \Pi\, k_j(p).$$

$$\overline{D(p)} = \frac{D(p)}{d_1(p)}.$$

Then

$$\Phi(0) = \frac{\overline{D(0)}}{D(0) + K(0)}\,k_1(0).$$

If $D(0) = 0$, then $\Phi(0)$ is also equal to zero. But $D(0) = 0$ only when the controller contains an astatic element.

Therefore, *a single-loop system is astatic if the controller contains an astatic element. In the contrary case, the system is static if the controlled object is static.*

8. Frequency Characteristics of a Linear Element and of the Linear Model of a System

When the properties of all the elements in a system are given by their equations of motion, the transfer function of the system is the starting point for subsequent calculations.

Often the processes taking place in the separate elements (for example, in the controlled object) have been inadequately studied, and the derivation of initial equations for these elements is difficult.

In such cases the calculation is based not on the equations of motion, but on the so-called *frequency characteristics* of the system. The advantage of the frequency response method is that the necessary characteristics can be constructed from the linearized equations of the separate elements, and that they can also be found experimentally for elements whose equations are not known. It is only necessary to ensure that the elements whose frequency characteristics have been determined experimentally were linear or nearly so. The methods for the experimental determination of frequency characteristics, described below, enable us at the same time to find out whether the given element is linear or not.

(a) *The frequency characteristics of a linear element*

We repeat the experiment described in Section 2, but instead of a sharp momentary change in the input coordinate of the given element we apply a sinusoidal signal of the form $X_{\text{in}} = A \sin \omega t$ at the input of the element. The generator producing this signal is so constructed that it permits one to change the frequency ω and the amplitude A of the applied disturbance within wide limits.

First let ω have some fixed value ω_1, and A be small within some limits. The disturbance in the input coordinate causes forced oscilla-

tory movements of the output coordinate. Let us record these using some oscillation recording device (oscillograph, vibrograph, etc.).

If the element is linear, then the output coordinate performs oscillations according to the law

$$X_{\text{out}} = B_1 \sin(\omega_1 t + \varphi_1)\,.$$

Let us draw the vector from the origin of the u, v plane whose modulus (the length of the vector) is equal to $r_1 = \dfrac{A}{B_1}$, and whose argument (the angle between the positive direction of the u-axis and the vector) is equal to the angle φ_1 with the opposite sign (usually φ_1 is negative). At the end of the vector we put the point ω_1 (Fig. 89).

We now change the frequency ω of the oscillations applied at the input of the system, without changing their amplitude and phase.

Let $\omega = \omega_2$. Then for oscillation of frequency ω_2 :

$$X_{\text{out}} = B_2 \sin(\omega_2 t + \varphi_2)\,.$$

In Fig. 89 we construct a vector with modulus $r_2 = A/B_2$, argument $-\varphi_2$ and end-point ω_2. Similarly, repeating the experiment for a new value ω_3 we construct in the same figure a vector with $\omega = \omega_3$, as end-point, and so on for various values of ω from $\omega = 0$ to $\omega = \infty$ (in practice, up to sufficiently large ω) and we join the ends of these vectors with a smooth curve.

The curve constructed in this way is called the *amplitude-phase characteristic of the linear element.*

In order to determine from the amplitude-phase characteristic of a linear element the amplitude and phase of the oscillations of the output coordinate, knowing the amplitude A and frequency of oscillation $\omega = \omega^*$ of the input coordinate, we must find that point on the amplitude-phase characteristic which has the end-point $\omega = \omega^*$. Let the vector produced from the origin of coordinates to this point have length r and argument φ. Then the amplitude of oscillations of X_{out} is equal to $B = \dfrac{A}{r}$ and the phase is equal to $-\varphi$ (Fig. 89).

The amplitude-phase characteristics are often constructed differently: the length of each vector is taken as being equal not to $\dfrac{A}{B}$,

but to $\frac{B}{A}$, and the argument not to $-\varphi$ but to φ. In order to distinguish these two methods of constructing the characteristics, we agree to call characteristics for which $r = \frac{A}{B}$ *amplitude-phase characteristics of the first kind*, and those for which $r = \frac{B}{A}$ characteristics of the

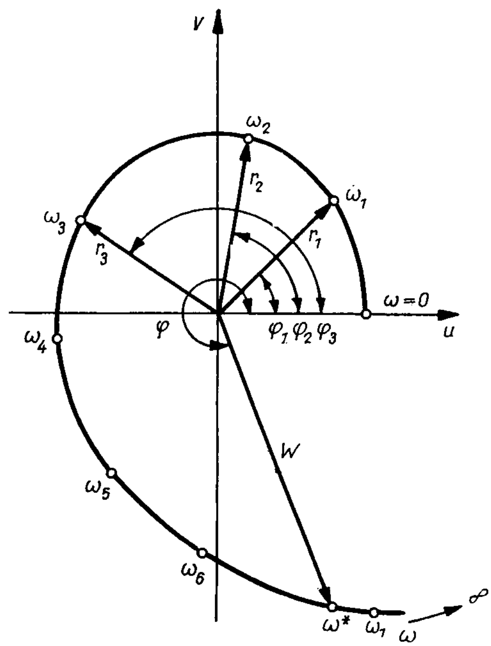


FIG. 89

second kind.* Figure 90 gives some examples of amplitude-phase characteristics of the first and second kinds for various typical stages.

The connexion between the amplitude-phase characteristics of the first and second kinds is established by the vector equation

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = 1,$$

* In the literature on the subject, amplitude-phase characteristics of the first kind are often called reverse or inverse, and of the second kind, usual or simple. In our view these terms are unfortunate, since the use of amplitude-phase characteristics of the first kind is often considerably more convenient and simple than that of the second kind.

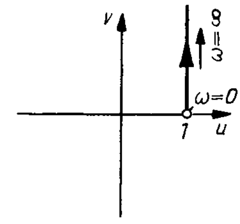
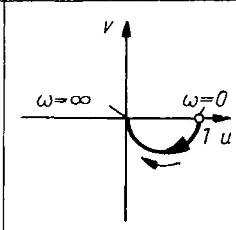
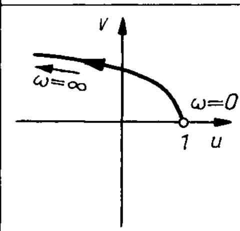
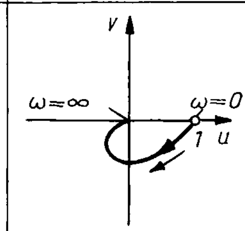
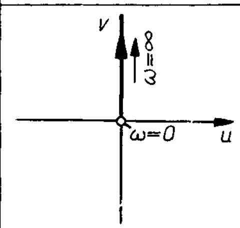
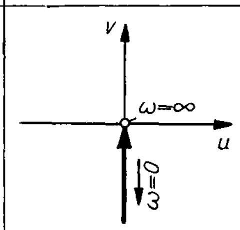
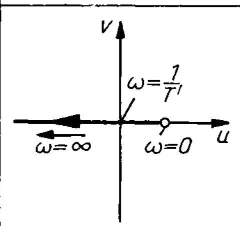
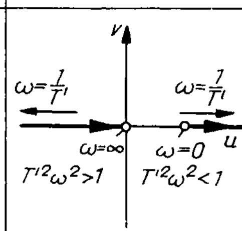
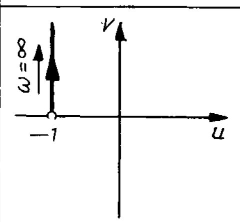
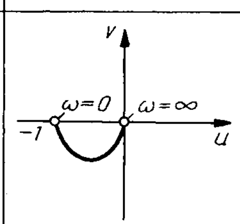
Name of stage	Stage operator	Amplitude-phase characteristic of the first kind	Amplitude-phase characteristic of the second time
Single capacitance	$Tp+1$		
Oscillatory	$T'^2p^2+T_kp+1$		
Astatic	p		
Conservative	T'^2p^2+1		
Unstable	$Tp-1$		

FIG. 90

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the vectors produced from the co-ordinate origin to the points with the same ω on the amplitude-phase characteristics of the first and second kind. The vectors here are multiplied according to the usual rules, i.e. their moduli are multiplied together, and their arguments are added.

The curve relating the value of $\frac{A}{B}\left(\text{or } \frac{B}{A}\right)$ to ω is called the *amplitude characteristic* of the element, and the curve relating the value of $-\varphi$ (or $+\varphi$) to ω its *phase characteristic.*

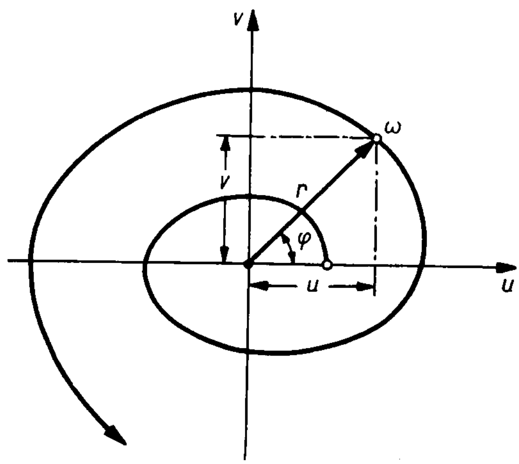


FIG. 91

Let us call u and v the projections of a vector of the amplitude-phase characteristic on the abscissa and ordinate axes (Fig. 91).

The curves relating u and v to ω are called respectively the *real* and *imaginary characteristics* of the element. The reason for these names will become clear in the following section.

The term *"frequency characteristics"* includes the characteristics of all five types enumerated above: amplitude-phase, amplitude, phase, real and imaginary.

For the linear element, the amplitude B is proportional to A and therefore the construction of the amplitude-phase characteristic is independent of A. Thus *an element is linear if, when a harmonic oscillation is applied to its input a harmonic oscillation is also set up at the output, with the same frequency as, and with an amplitude proportional to the amplitude of the input oscillation.*

If an element is not linear, then the shape of the oscillations of the output coordinate is not sinusoidal and its ordinates are not proportional to A. The oscillations of the output coordinate in this case may be represented by a Fourier series and after this we may construct the amplitude-phase characteristics separately for each harmonic. But to do this it is necessary to construct a separate curve for each value of A.

If the element is non-linear, but is linearizable, then the output oscillations depend also on the position of the mean (equilibrium) value of the input coordinate in the non-linear static characteristic of the element.

In this case, in order to obtain correctly the amplitude-phase characteristic of the element experimentally, it is necessary to apply sinusoidal oscillations of small amplitude to the input coordinate. These oscillations must depend on the value of the input coordinate which is steady in the conditions being considered. This value is found from the static solution and was taken earlier as the zero reading in the derivation of the equations. Each set of conditions has its own amplitude-phase characteristic, so that for each set it is necessary to carry out a series of experiments, gradually increasing the amplitude of the applied oscillations until such time as a change in this amplitude does not any longer give rise to a change in the curve of the amplitude-phase characteristic obtained. Thus, for each set of conditions we must determine the amplitude-phase characteristic and define the region in which a linear analysis is valid.

Experience has shown that the method of collecting such a series of frequency characteristics for linearizable elements is little used and the experimental determination of the frequency characteristics is widely adopted only for linear elements.

(b) *The determination of the amplitude-phase characteristic of a linear model from the equations of motion*

When equations of motion have been derived for some elements of the system, but the properties of other, linear, elements are given by experimentally derived frequency characteristics, it is necessary to construct the frequency characteristics for the first elements*

* The converse problem of determining the equations of motion from the frequency characteristics is considerably more complicated, and in practice, as far as we can, we avoid it.

from their equations of motion, in order to construct the frequency characteristic of the system from the frequency characteristics of all the elements and to take it as the initial material for further calculations.

We use Euler's identity :

$$e^{i\omega t} = \cos \omega t + i \sin \omega t .$$

If we replace the disturbance which we applied at the input of the element when we obtained the frequency characteristic

$$X_{\text{in}} = A \sin \omega t ,$$

by the complex function

$$X_{\text{in}}^{*} = A e^{i\omega t} , \tag{2.47}$$

then the true disturbance is the imaginary part of (2.47).

Putting this value of X_{in}^{*} into the equation of the element, we calculate the forced motions of the output coordinate by finding the particular integral of this equation.

Because of the principle of superposition, its imaginary part also defines forced motions of the output coordinate of the open system, caused by an external disturbance

$$X_{\text{in}} = A \cdot \sin \omega t .$$

Let the most general form of the equation of the element be

$$\begin{aligned} a_0 \frac{d^n X_{\text{out}}}{dt^n} + a_1 \frac{d^{n-1} X_{\text{out}}}{dt^{n-1}} + \ldots + a_n X_{\text{out}} = \\ = b_0 \frac{d^m X_{\text{lin}}}{dt^m} + b_1 \frac{d^{m-1} X_{\text{lin}}}{dt^{m-1}} + \ldots + b_m X_{\text{lin}} . \end{aligned} \tag{2.48}$$

We will look for the particular integral of (2.48) in the form

$$X_{\text{out}} = B e^{i(\omega t + \varphi)} = B e^{i\omega t} e^{i\varphi} .$$

We note that

$$\frac{dX_{out}}{dt} = Be^{i\varphi}(i\omega)\, e^{i\omega t},$$

$$\frac{d^2 X_{out}}{dt^2} = Be^{i\varphi}(i\omega)^2\, e^{i\omega t}\,.$$

$$\cdots\cdots\cdots\cdots$$

$$\frac{d^n X_{out}}{dt^n} = Be^{i\varphi}(i\omega)^n\, e^{i\omega t}\,.$$

Substituting these values in (2.48) we find

$$\begin{aligned}[a_0(i\omega)^n + a_1(i\omega)^{n-1} + \ldots + a_n]\, Be^{i\varphi} e^{i\omega t} = \\ = [b_2(i\omega)^m + b_1(i\omega)^{m-1} + \ldots + b_m]\, Ae^{i\omega t}\end{aligned} \tag{2.49}$$

or

$$d(i\omega)\cdot Be^{i\varphi} e^{i\omega t} = k(i\omega)\, Ae^{i\omega t}.$$

Hence

$$\frac{A}{B} e^{-i\varphi} = \frac{d(i\omega)}{k(i\omega)}\,. \tag{2.50}$$

Separating real and imaginary parts in (2.50) we can write it as :

$$\frac{A}{B} e^{-i\varphi} = u(\omega) + iv(\omega) = \sqrt{[u(\omega)]^2 + [v(\omega)]^2}\, e^{i \arctan \frac{v(\omega)}{u(\omega)}}\,.$$

Hence

$$\frac{A}{B} = \sqrt{[u(\omega)]^2 + [v(\omega)]^2} \tag{2.51}$$

and

$$\varphi = -\arctan \frac{v(\omega)}{u(\omega)}\,. \tag{2.52}$$

Thus, for an external disturbance

$$X_{in} = Ae^{i\omega t}$$

the forced motions of the output coordinate are equal to

$$X_{out} = \frac{A}{\sqrt{[u(\omega)]^2 + [v(\omega)]^2}}\, e^{\left[i\left(\omega t - \arctan \frac{v(\omega)}{u(\omega)}\right)\right]}$$

or

$$X_{\text{out}} = \frac{A}{\sqrt{[u(\omega)]^2 + [v(\omega)]^2}} \left\{ \cos\left[\omega t - \arctan\frac{v(\omega)}{u(\omega)}\right] + \right.$$
$$\left. + i \sin\left[\omega t - \arctan\frac{v(\omega)}{u(\omega)}\right]\right\}. \tag{2.53}$$

The actual external disturbance has the form*

$$X_{\text{in}} = A \sin \omega t = \text{Im}\, A e^{i\omega t}.$$

Thus the oscillations at the output are determined by the imaginary part of (2.53), i.e.

$$X_{\text{out}} = \frac{A}{\sqrt{[u(\omega)]^2 + [v(\omega)]^2}} \sin\left[\omega t - \text{arc tan}\,\frac{v(\omega)}{u(\omega)}\right].$$

The length of the vector from the origin to a point of the amplitude-phase characteristic of the first kind is determined by the formula (2.51) and its argument by (2.52).

The formula (2.51) determines the amplitude, and (2.52) the phase of the characteristic. The expressions for $u(\omega)$ and $v(\omega)$ define the real and imaginary characteristics.

Knowing the equations of motion, i.e. $d(p)$ and $k(p)$, it is easy to find from these formulae a point of the amplitude-phase characteristic for any ω.

(c) *The construction of the frequency characteristic of an element from its time characteristic*

Frequently the properties of some elements of a system are given by their amplitude-phase characteristics and the properties of its other elements are given by their time characteristics, since sometimes an element does not allow the application of oscillations, while the experimental determination of the time characteristic does not constitute any difficulty (see Section 1 of this chapter).

* Re denotes the real part, and Im the coefficient of the imaginary part of the expression following these letters.

For example, if $W(i\omega) = u(\omega) + iv(\omega)$, then $\text{Re}\, W(i\omega) = u(\omega)$ and $\text{Im}\, W(i\omega) = v(\omega)$.

In such cases it is necessary to construct the frequency characteristic of the element from its time characteristic.

Let the time characteristic of an element be given (as in Fig. 92).

We assume that we have introduced relative coordinates, and that the element receives at its input a disturbance equal to 0 for $t < 0$ and equal to 1 for all $t \geqslant 0$ in the given relative coordinate system.

We divide the time axis (t-axis) into m small equal intervals of time Δt. According to the time characteristic, each such interval

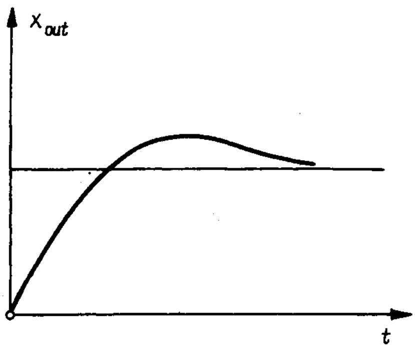


FIG. 92

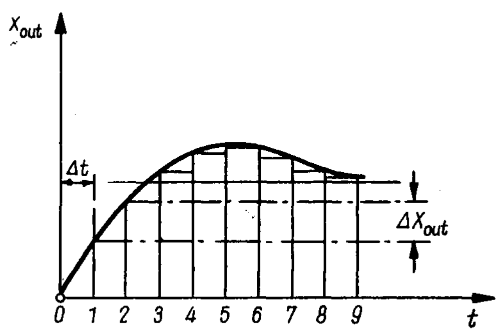


FIG. 93

corresponds to a definite positive or negative increment in the output coordinate Δx_{out} (Fig. 93).

Let us set ω equal to any value ω_1, and construct m vectors at the origin in the (u, v)-plane in such a way that the modulus of the jth vector is equal to Δx_j, the first of the vectors is directed along the u-axis, and the increment in the argument of the j-th vector with respect to the argument of the preceding $(j - 1)$-th vector, reading clockwise, is equal to $\omega_1 \Delta t$. The increment in the argument is decreased by an additional 180° for those vectors which correspond to the point where Δx_j changes sign. The sum of all these vectors determines a vector of the amplitude-phase characteristic of the second kind of the element we are considering for the value ω_1.

It is then necessary to repeat this construction with new values of ω, and to determine the characteristic vector for these values.

We demonstrate this construction (Fig. 94) using the time characteristic shown in Fig. 93, putting $\omega_1 = 1$.

We construct the vector of modulus Δx_1 and argument zero. At the end of this vector we construct a new vector of modulus Δx_2 and argument, reading clockwise from the direction of the first vector, equal to Δt. From the end of this vector we draw a vector with modulus Δx_3, and argument Δt, reading clockwise from the direction of the second vector, and so on.

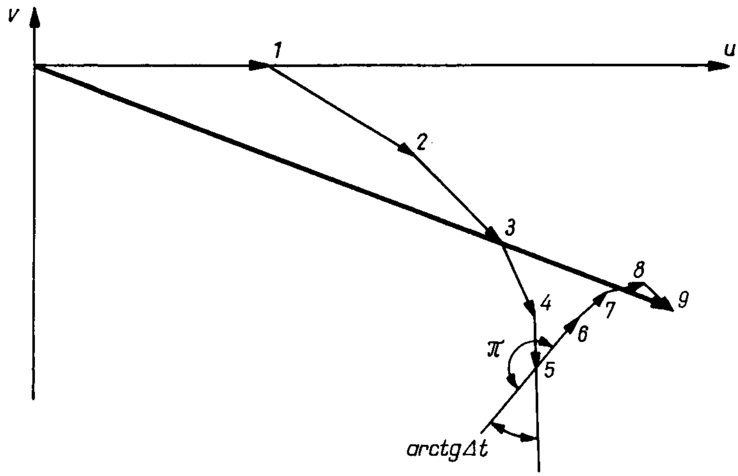


FIG. 94

Starting from the fifth vector (Δx_5 and further) the increments Δx are negative (see Fig. 93), and therefore the increment of the sixth vector with respect to the fifth is equal to $\Delta t - \pi$.

The vector joining the origin of co-ordinates to the end of the m-th vector of this construction is also the characteristic vector for frequency $\omega = 1$. This means that the end of the m-th vector determines the point of the amplitude-phase characteristic of the second kind corresponding to the frequency $\omega = 1$.

The greater the value of m, and the smaller the value of Δt, the higher is the accuracy of the result. When the values of ω can be bounded so that $2\omega\, \Delta t < 1$, the construction can be made more accurate by multiplying the modulus of the resulting vector obtained for any value $\omega = \omega_j$ by the quantity

$$\frac{\pi\omega_j \Delta t}{\sin(\pi\omega_j \Delta t)} \cdot$$

The basic advantage of this method for constructing the amplitude-phase characteristic from the time characteristic is that the whole construction is done graphically and does not demand the use of tables or any calculation : the moduli of the vectors are found directly from Fig. 93 and the increments in the arguments of all the summed vectors are identical. The disadvantage of this method lies in the fact that in order to obtain one point of the amplitude-phase characteristic a large number (often several dozen) of vectors must be constructed. It is possible to reduce the number of vectors summed

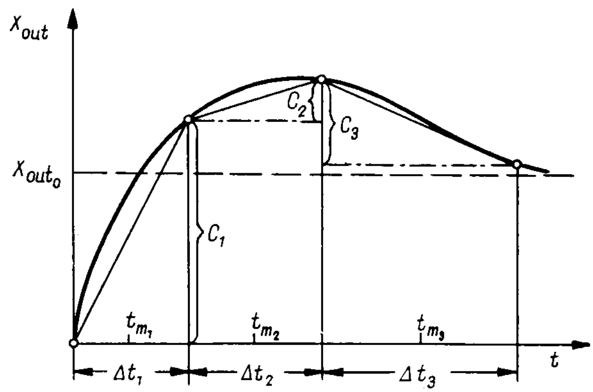


FIG. 95

by many times at the cost of some complication in computing their moduli and arguments, by using a second construction arising from the equation

$$f(i\omega) = \sum_{j=1}^{j=r} \left[\frac{2C_j}{\Delta t_j \omega} \sin \frac{\Delta t_j \omega}{2} \right] e^{-i t_{m_j} \omega}.$$

This equation is easily derived from the basic theory of Laplace transforms. In order to use it we must replace the time characteristic not by the stepped line but by a series of straight line segments (Fig. 95) and drop perpendiculars from the points of intersection of adjacent segments on to the t-axis. These perpendiculars divide the t-axis into r sections. The lengths of these sections are denoted by Δt_j, the abscissa corresponding to the mid-point of each section is t_{m_j}, and the projections of the broken line on the ordinate axis are C_j.

The number of vectors which we must sum in this method in order to obtain one point of the amplitude-phase characteristic is equal to the number of straight line segments which make up the broken line in Fig. 95. This is considerably smaller than the number of steps in Fig. 93. But on the other hand the modulus and argument of each vector must be calculated from the given formula.

(d) *The construction of the frequency characteristic of a system from the frequency characteristics of its elements*

Let the Laplace transform of the input and output coordinates of a linear element (or system) be connected by the relation

$$M(p)\, L[x_{\text{out}}] = R(p)\, L[x_{\text{in}}],$$

where $M(p)$ and $R(p)$ are polynomials in p, so that the transfer function of this element (or system) is

$$W(p) = \frac{R(p)}{M(p)}.$$

Then, as we showed above, the amplitude-phase characteristic of the second kind for this element (or system) can be constructed if the ratio

$$\frac{R(i\omega)}{M(i\omega)} = u(\omega) + iv(\omega),$$

is found for a range of values of ω say from 0 to $+\infty$ and the ends of the vectors corresponding to these complex numbers, joined by a smooth curve. Thus, the amplitude-phase characteristic of the second kind of an element (part of a system or a system as a whole) can be obtained by substituting $i\omega$ for p in its transfer function.

In Section 6 it was shown how the transfer function of the open or closed system is related to the separate transfer functions of its elements.

These same relationships, after substituting $i\omega$ for p, define the relation between the amplitude-phase characteristics of the second kind of an open or closed system and the amplitude-phase characteristics of the second kind of its elements.

Thus, for example, from (2.36) it follows that each vector of the amplitude-phase characteristic of the second kind of an open sequen-

tial circuit of elements is found for any value of ω by multiplying* the vectors of the amplitude-phase characteristics of the second kind of all the elements for this value of ω.

If for each element

$$w_j(i\omega) = r_j e^{i\varphi_j(\omega)},$$

then for a sequential circuit of n elements

$$W(i\omega) = \prod_{j=1}^{j=n} w_j(i\omega) = \prod_{j=1}^{j=n} r_j e^{i\varphi_j(\omega)} = r(\omega) e^{i\varphi(\omega)},$$

where

$$r(\omega) = \prod_{j=1}^{j=n} r_j(\omega) \text{ and } \varphi(\omega) = \sum_{j=1}^{j=n} \varphi_j(\omega).$$

For a second example we consider a closed single-loop circuit.

If an external disturbance is applied to the first element, then the transfer function for this disturbance of the coordinate x_K is determined by the formula

$$\Phi(p) = \frac{W_1(p)}{1 + W_1(p) W_2(p)}$$

If we assume that the external disturbance is sinusoidal, then the steady oscillations of the coordinate x_K are determined by the amplitude-phase characteristic

$$\Phi(i\omega) = \frac{W_1(i\omega)}{1 + W_1(i\omega) W_2(i\omega)},$$

where

$$W_1(i\omega) = \prod_{j=1}^{j=k} w_j(i\omega) \text{ and } W_2(i\omega) = \prod_{j=k+1}^{j=n} w_j(i\omega).$$

In order to construct it, we must first construct the hodograph of $W_1(i\omega)$ and the hodograph of $W_1(i\omega) \cdot W_2(i\omega)$. The imaginary axis on the latter graph must then be displaced to the left by one unit. In this way we obtain the hodograph of

$$1 + W_1(i\omega) W_2(i\omega).$$

Then, the vectors $W_1(i\omega)$ must be divided by the vectors of the constructed hodograph $1 + W_1(i\omega) W_2(i\omega)$ for the same value of ω.

* Here, vectors are multiplied according to the multiplication rules for complex numbers, the moduli being multiplied, the arguments added.

Just as the formation of the transfer function of the system completes the preparation for the analysis of the control process using the equations of the element, so the formation of the amplitude-phase characteristic of the system completes the preparation for the analysis of the control process using the amplitude-phase characteristics of its elements.

To analyse the process, the characteristics both of the open system and of the closed system must be available. (For an investigation of stability see Chapter III, and for a study of the course of the process, see Chapter IV.)

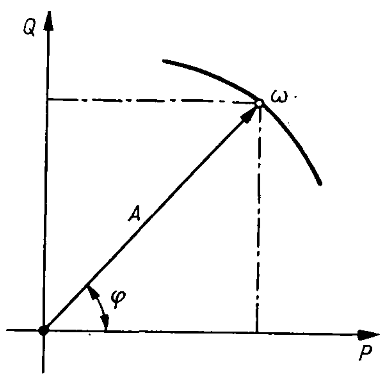


FIG. 96

For an open system it is essential to know the amplitude-phase characteristic, while for a closed system we are usually required to know the amplitude and the real characteristics. Let us represent the vector $\Phi(i\omega)$ of the closed system in the two forms :

$$\Phi(i\omega) = A(\omega)\, e^{i\varphi(\omega)} \text{ and } \Phi(i\omega) = P(\omega) + iQ(\omega).$$

The graph of the function $A(\omega)$ is called the amplitude characteristic of the closed system, and the graph of the function $P(\omega)$ is its real characteristic.

In Fig. 96 are shown the amplitude-phase characteristic, the vector $\Phi(i\omega)$ (for one value of ω), and the quantities P, Q, A and φ are indicated. From Fig. 96 it follows that

$$A^2(\omega) = P^2(\omega) + Q^2(\omega)$$

and

$$P(\omega) = A(\omega) \cdot \cos\varphi(\omega).$$

It is usually sufficient to carry out the construction of the amplitude-phase characteristic of the open system ; from it, it is then easy to construct the amplitude characteristic $A(\omega)$ and the real characteristic $P(\omega)$ for the closed system.

(e) *Logarithmic characteristics*

When the transfer function of the open system is equal to the product of the transfer functions of the separate elements, the construction of the amplitude-phase characteristic of the system from the characteristics of the elements is simplified if a logarithmic scale is employed.*

The transition to a logarithmic scale also extremely simplifies the construction of the amplitude-phase characteristics of the elements from their linear equations, in particular in cases where the degree of the operators $d_j(p)$ and $k_j(p)$ is not greater than one.

Terminology. The terminology used in the construction of logarithmic frequency characteristics is borrowed from acoustics.

If two frequencies are such that one is twice the other, that is

$$\frac{\omega_2}{\omega_1} = 2,$$

then the frequencies ω_1 and ω_2 are said to differ from one another by one *octave.*

If this ratio is equal to ten, i. e.

$$\frac{\omega_2}{\omega_1} = 10,$$

then these frequencies are said to differ by one *decade.*

In order to measure the ratio of two quantities which vary over a wide range, the logarithmic scale is often used.

In measuring the ratio of two powers N_1 and N_2, they are said to differ by one *bel* if

$$\log \frac{N_2}{N_1} = 1.$$

This is a comparatively large unit of measurement.

* Various methods have been suggested which use the logarithmic scale, for the construction of the frequency characteristics of non-single-loop systems as well, but for these the transition to logarithmic characteristics still has not such undeniable advantages as for single-loop systems.

When considering concrete problems it is usually necessary to use a smaller unit of measurement, called the *decibel*. This unit is defined by the following equation :

$$10 \log \frac{N_2}{N_1} = 1.$$

From powers, we may pass to the measurement of "protoplastic" quantities (the amplitudes of the forces, of current, potential, pressure and so on) whose square is proportional to the powers

$$N_1 \equiv J_1^2, \; N_2 \equiv J_2^2.$$

Thus, if

$$10 \log \frac{N_2}{N_1} = 10 \log \frac{J_2^2}{J_1^2} = 1,$$

then the ratio of the amplitudes of the forces of current, potential, pressure and so forth is given by

$$20 \log \frac{J_2}{J_1} = 1$$

if J_2 differs from J_1 by 1 decibel.

Table I enables us to translate logarithmic units of measurement into units of the ratio of powers and of "protoplastic" quantities (e.g. potentials or currents).

Henceforth, when we use logarithmic characteristics the frequency measurements will be made in octaves or decades, and the amplitude in decibels. Proceeding to the construction of logarithmic characteristics, we begin by considering an astatic stage.

Astatic stage. The transfer function of this stage is defined by the equation

$$w(p) = \frac{k}{p}.$$

Putting $p = i\omega$, in this equation we obtain :

$$w(i\omega) = \frac{k}{i\omega}. \tag{2.54}$$

TABLE I.

CONVERSION FROM DECIBELS INTO ABSOLUTE RATIOS

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
1·0000	1·0000	0	1·0000	1·0000
0·9994	0·9989	0·005	1·0006	1·0012
0·9989	0·9977	0·01	1·0012	1·0023
0·9977	0·9954	0·02	1·0023	1·0046
0·9966	0·9931	0·03	1·0035	1·0069
0·9954	0·9908	0·04	1·0046	1·0093
0·9942	0·9886	0·05	1·0058	1·0116
0·9931	0·9863	0·06	1·0069	1·0139
0·9920	0·9840	0·07	1·0081	1·0162
0·9908	0·9818	0·08	1·0093	1·0186
0·9897	0·9704	0·09	1·0104	1·0209
0·9886	0·9772	0·1	1·012	1·023
0·9772	0·9550	0·2	1·023	1·047
0·9661	0·9333	0·3	1·035	1·072
0·9550	0·9120	0·4	1·047	1·096
0·9441	0·8913	0·5	1·059	1·122
0·9333	0·8710	0·6	1·072	1·148
0·9226	0·8511	0·7	1·084	1·175
0·9120	0·8318	0·8	1·096	1·202
0·9016	0·8128	0·9	1·109	1·230
0·8913	**0·7943**	**1·0**	**1·122**	**1·259**
0·8810	0·7762	1·1	1·135	1·288
0·8710	0·7586	1·2	1·148	1·318
0·8610	0·7413	1·3	1·161	1·349
0·8511	0·7244	1·4	1·175	1·380
0·8414	0·7079	1·5	1·189	1·413
0·8318	0·6918	1·6	1·202	1·445
0·8222	0·6761	1·7	1·216	1·479
0·8128	0·6607	1·8	1·230	1·514
0·8035	0·6457	1·9	1·245	1·549
0·7943	**0·6310**	**2·0**	**1·259**	**1·585**
0·7852	0·6166	2·1	1·274	1·622
0·7762	0·6026	2·2	1·288	1·660
0·7674	0·5888	2·3	1·303	1·698
0·7586	0·5754	2·4	1·318	1·738

TABLE I. (contd.)

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
0·7499	0·5623	2·5	1·334	1·778
0·7413	0·5495	2·6	1·349	1·820
0·7328	0·5370	2·7	1·365	1·862
0·7244	0·5248	2·8	1·380	1·905
0·7161	0·5129	2·9	1·396	1·950
0·7079	**0·5012**	**3·0**	**1·413**	**1·995**
0·6998	0·4898	3·1	1·429	2·042
0·6918	0·4786	3·2	1·445	2·089
0·6839	0·4677	3·3	1·462	2·138
0·6761	0·4571	3·4	1·479	2·188
0·6683	0·4467	3·5	1·496	2·239
0·6607	0·4365	3·6	1·514	2·291
0·6531	0·4266	3·7	1·531	2·344
0·6457	0·4169	3·8	1·549	2·399
0·6383	0·4074	3·9	1·567	2·455
0·6310	**0·3981**	**4·0**	**1·585**	**2·512**
0·6237	0·3890	4·1	1·603	2·570
0·6166	0·3802	4·2	1·622	2·630
0·6095	0·3715	4·3	1·641	2·692
0·6026	0·3631	4·4	1·660	2·754
0·5957	0·3548	4·5	1·679	2·818
0·5888	0·3467	4·6	1·698	2·884
0·5821	0·3388	4·7	1·718	2·951
0·5754	0·3311	4·8	1·738	3·020
0·5689	0·3336	4·9	1·758	3·090
0·5623	**0·3162**	**5·0**	**1·778**	**3·162**
0·5559	0·3090	5·1	1·799	3·236
0·5495	0,3020	5·2	1·820	3·311
0·5433	0·2951	5·3	1·841	3·388
0·5370	0·2884	5·4	1·862	3·467
0·5309	0·2818	5·5	1·884	3·548
0·5248	0·2754	5·6	1·905	3·631
0·5188	0·2692	5·7	1·928	3·715
0·5129	0·2630	5·8	1·950	3·802
0·5070	0·2570	5·9	1·972	3·890

TABLE I. (contd.)

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
0·5012	**0·2512**	**6·0**	**1·995**	**3·981**
0·4955	0·2455	6·1	2·018	4·074
0·4898	0·2399	6·2	2·042	4·169
0·4842	0·2444	6·3	2·065	4·266
0·4786	0·2291	6·4	2·089	4·366
0·4732	0·2239	6·5	2·113	4·467
0·4677	0·2188	6·6	2·138	4·571
0·4624	0·2138	6·7	2·163	4·677
0·4571	0·2089	6·8	2·188	4·786
0·4519	0·2042	6·9	2·213	4·898
0·4467	**0·1995**	**7·0**	**2·239**	**5·012**
0·4416	0·1950	7·1	2·265	5·129
0·4365	0·1905	7·2	2·291	5·248
0·4315	0·1862	7·3	2·317	5·370
0·4266	0·1820	7·4	2·344	5·495
0·4217	0·1778	7·5	2·371	5·623
0·4169	0·1738	7·6	2·399	5·754
0·4121	0·1698	7·7	2·427	5·888
0·4074	0·1660	7·8	2·455	6·026
0·4027	0·1622	7·9	2·483	6·166
0·3981	**0·1585**	**8·0**	**2·512**	**6·310**
0·3936	0·1349	8·1	2·541	6·457
0·3890	0·1514	8·2	2·570	6·607
0·3846	0·1479	8·3	2·600	6·761
0·3802	0·1445	8·4	2·630	6·918
0·3758	0·1413	8·5	2·661	7·079
0·3715	0·1380	8·6	2·692	7·244
0·3673	0·1349	8·7	2·723	7·413
0·3631	0·1318	8·8	2·754	7·586
0·3589	0·1288	8·9	2·786	7·762
0·3548	**0·1259**	**9·0**	**2·818**	**7·943**
0·3508	0·1230	9·1	2·851	8·128
0·3467	0·1202	9·2	1·884	8·318
0·3428	0·1175	9·3	2·917	8·511
0·3388	0·1148	9·4	2·951	8·710

TABLE I. (contd.)

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
0·3350	0·1122	9·5	2·985	8·913
0·3311	0·1096	9·6	3·020	9·120
0·3273	0·1072	9·7	3·055	9·333
0·3236	0·1047	9·8	3·090	9·550
0·3199	0·1023	9·9	3·126	9·772
0·3162	**0·10000**	**10·0**	**3·162**	**10·00**
0·3126	0·09772	10·1	3·199	10·23
0·3090	0·09550	10·2	3·236	1·047
0·3055	0·09333	10·3	3·273	10·72
0·3020	0·91200	10·4	3·311	10·66
0·2985	0·08913	10·5	3·350	11·22
0·2951	0·08710	10·6	3·388	11·48
0·2917	0·08511	10·7	3·428	11·75
0·2884	0·08318	10·8	3·467	12·02
0·2851	0·08128	10·9	3·508	12·30
0·2818	**0·07943**	**11·0**	**3·548**	**12·59**
0·2786	0·07762	11·1	3·589	12·88
0·2754	0·07586	11·2	3·631	13·18
0·2723	0·07413	11·3	3·673	13·49
0·2692	0·07244	11·4	3·715	13·80
0·2661	0·07079	11·5	3·758	14·13
0·2630	0·06918	11·6	3·802	14·45
0·2600	0·06761	11·7	3·846	14·79
0·2570	0·06607	11·8	3·890	15·14
0·2541	0·06457	11·9	3·936	15·49
0·2512	**0·06310**	**12·0**	**3·981**	**15·85**
0·2483	0·06166	12·1	4·027	16·22
0·2455	0·06026	12·2	4·074	16·60
0·2427	0·05888	12·3	4·121	16·98
0·2399	0·05754	12·4	4·169	17·38
0·2371	0·05623	12·5	4·271	17·78
0·2344	0·05495	12·6	4·266	18·20
0·2317	0·05370	12·7	4·315	18·62
0·2291	0·05248	12·8	4·365	19·05
0·2265	0·05129	12·9	4·416	19·50

TABLE I (contd.)

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
0·2239	**0·05012**	**13·0**	**4·467**	**19·95**
0·2213	0·04898	13·1	4·519	20·42
0·2188	0·04786	13·2	4·571	20·89
0·2163	0·04677	13·3	4·624	21·38
0·2138	0·04571	13·4	4·677	21·88
0·2113	0·04467	13·5	4·732	22·39
0·2089	0·04365	13·6	4·786	22·91
0·2065	0·04266	13·7	4·842	23·44
0·2042	0·04169	13·8	4·898	23·99
0·2018	0·04074	13·9	4·955	24·55
0·1995	**0·03981**	**14·0**	**5·012**	**25·12**
0·1972	0·03890	14·1	5·070	25·70
0·1950	0·03802	14·2	5·129	26·30
0·1928	0·03715	14·3	5·188	26·92
0·1905	0·03631	14·4	5·248	27·54
0·1884	0·03548	14·5	5·309	28·18
0·1862	0·03567	14·6	5·370	28·84
0·1841	0·03488	14·7	5·433	29·51
0·1820	0·03311	14·8	5·495	30·20
0·1799	0·03236	14·9	5·559	30·90
0·1778	**0·03162**	**15·0**	**5·623**	**31·62**
0·1758	0·03090	15·1	5·689	32·36
0·1738	0·03020	15·2	5·754	33·11
0·1718	0·02951	15·3	5·821	33·83
0·1698	0·02884	15·4	5·888	34·67
0·1679	0·02818	15·5	5·957	35·48
0·1660	0·02754	15·6	6·026	36·31
0·1641	0·02692	15·7	6·095	37·15
0·1622	0·02630	15·8	6·166	38,02
0·1603	0·02470	15·9	6·237	38·90
0·1585	**0·02512**	**16·0**	**6·310**	**39·81**
0·1567	0·02455	16·1	6·383	40·74
0·1549	0·02399	16·2	6·457	41·69
0·1531	0·02344	16·3	6·531	42·66
0·1514	0·02291	16·4	6·607	43·65

TABLE I. (contd.)

Ratio of voltages or currents	Ratio of powers	±dB	Ratio of voltages or currents	Ratio of powers
0·1496	0·02239	16·5	6·683	44·67
0·1479	0·02188	16·6	6·761	45·71
0·1462	0·02138	16·7	6·839	46·77
0·1445	0·02089	16·8	6·918	47·86
0·1429	0·02042	16·9	6·998	48·98
0·1413	**0·01995**	**17·0**	**7·079**	**50·13**
0·1396	0·01950	17·1	7·161	51·29
0·1380	0·01905	17·2	7·244	52·48
0·1365	0·01862	17·3	7·328	53·70
0·1349	0·01820	17·4	7·413	54·95
0·1334	0·01778	17·5	7·499	56·23
0·1318	0·01738	17·6	7·586	57·54
0·1303	0·01698	17·7	7·674	58·88
0·1288	0·01660	17·8	7·762	60·26
0·1274	0·01622	17·9	7·852	61·66
0·1259	**0·01585**	**18·0**	**7·943**	**63·10**
0·1245	0·01549	18·1	8·035	64·57
0·1230	0·01514	18·2	8·128	66·07
0·1216	0·01479	18·3	8·222	77·61
0·1202	0·01445	18·4	8·318	69·18
0·1189	0·01413	18·5	8·414	70·79
0·1175	0·01380	18·6	8·511	72·44
0·1161	0·01349	18·7	8·610	74·13
0·1148	0·01318	18·8	8·710	75,86
0·1135	0·01288	18·9	8·811	77·62
0·1122	**0·01259**	**19·0**	**8·913**	**79·43**
0·1109	0·01230	19·1	9·016	81·28
0·1096	0·01202	19·2	9·120	83·18
0·1084	0·01175	19·3	9·226	85·11
0·1072	0·01148	19·4	9·333	87·10
0·1059	0·01122	19·5	9·441	89·13
0·1047	0·01096	19·6	9·550	91·20
0·1035	0·01072	19·7	9·661	93·33
0·1023	0·01047	19·8	9·772	95·50
0·1012	0·01023	19·9	9·886	97·72
0·1000	**0·01000**	**20·0**	**10·000**	**100·00**

The amplitude characteristic is determined by the modulus of the transfer function

$$A(\omega) = |w(i\omega)| = \frac{k}{\omega}. \tag{2.55}$$

Thus, the amplitude characteristic, measured in decibels, is defined by the following equations:

$$L = 20 \log A(w) = 20 \log \frac{k}{\omega},$$

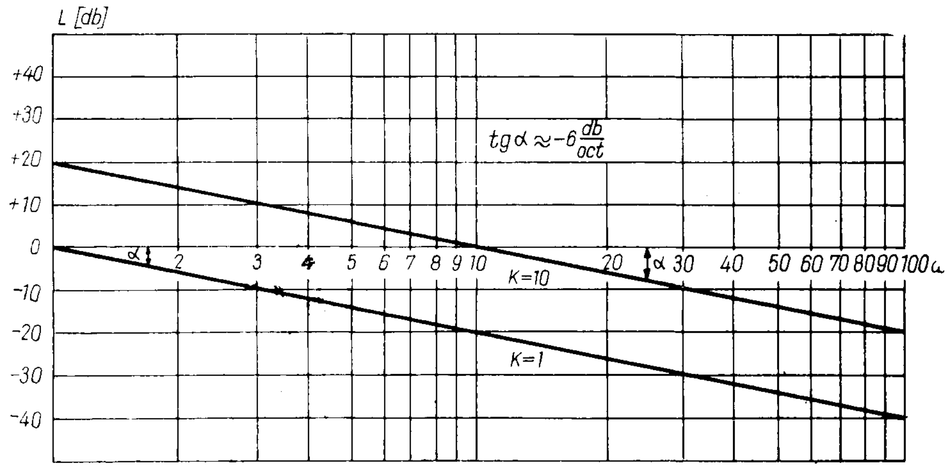


FIG. 97

or

$$L = 20 \lg k - 20 \lg \omega. \tag{2.56}$$

Let us put L, in decibels, along the y-axis, and $\log \omega$ along the x-axis, while indicating ω (Fig. 97). In these coordinates, equation (2,56) is the equation of a straight line. Usually the construction of a logarithmic amplitude characteristic is done first for $k = 1$, but the fact that $k \neq 1$ is taken into account later when passing from the characteristic of the element to that of the system.

For $k = 1$ from (2.56) we obtain $L = -20 \lg \omega$. When $\omega = \omega_1$ let $L = L_1$ and when $\omega = \omega_2 = 2\omega_1$ let $L = L_2$, where $L_1 = -20 \lg \omega_1$ $L_2 = -20 \lg \omega_2 = -20 \lg 2\omega_1 = -20 \lg \omega_1 - 20 \lg 2$. The latter equation can be written in the following form:

$$L_2 = L_1 - 20 \log 2 \approx L_1 - 6,$$

from which it follows that the straight line has a negative slope of 6 dB per octave.

Thus, the straight line with a negative slope of 6 d*B per octave passing through the origin of the coordinates* $\log \omega = 0$ $L = 0$ *i. e.* $\omega = 1$, *is the amplitude logarithmic characteristic of an astatic stage with coefficient of amplification* $k = 1$ *in coordinates* L, ω *(*L *in decibels and the frequency scale being logarithmic).*

Single-capacitance stage. The transfer function of a single-capacitance stage is

$$w(p) = \frac{k}{Tp + 1}.$$

If we put $\omega_T = \frac{1}{T}$, then

$$w(p) = \frac{k\omega_T}{p + \omega_T},$$

where ω_T is the *conditional* or *conjugate** frequency of the single-capacitance stage. Hence

* The reasons for this name will be explained below.

$$w(i\omega) = \frac{k\omega_T}{i\omega + \omega_T}.$$

The amplitude characteristic is given by the modulus of the transfer function

$$A(\omega) = \frac{k\omega_T}{\sqrt{\omega_T^2 + \omega^2}}.$$

As before, $L = 20 \log A(\omega)$, i. e.

$$L = 20 \log k + 20 \log \omega_T - 20 \log \sqrt{\omega_T^2 + \omega^2}.$$

If $k = 1$, then the equation of the characteristic may be written in the form

$$L = 20 \log \omega_T - 20 \log \sqrt{\omega_T^2 + \omega^2}.$$

We can considerably simplify the construction of this curve by finding its asymptotes.

Let $\omega \to 0$. Then $L = 20 \log \omega_T - 20 \log \sqrt{\omega_T^2 + \omega^2} \to 0$ and by putting an equals sign in place of $\to$, we obtain the equation of one of the asymptotes of the required curve.

If we let $\omega \to \infty$, then

$$\sqrt{\omega_T^2 + \omega^2} \to \sqrt{\omega^2} \text{ and } L \to 20 \log \omega_T - 20 \log \omega .$$

By putting $=$ instead of $\to$ we find the equation of the second asymptote to the required curve. It is clear that this asymptote passes through

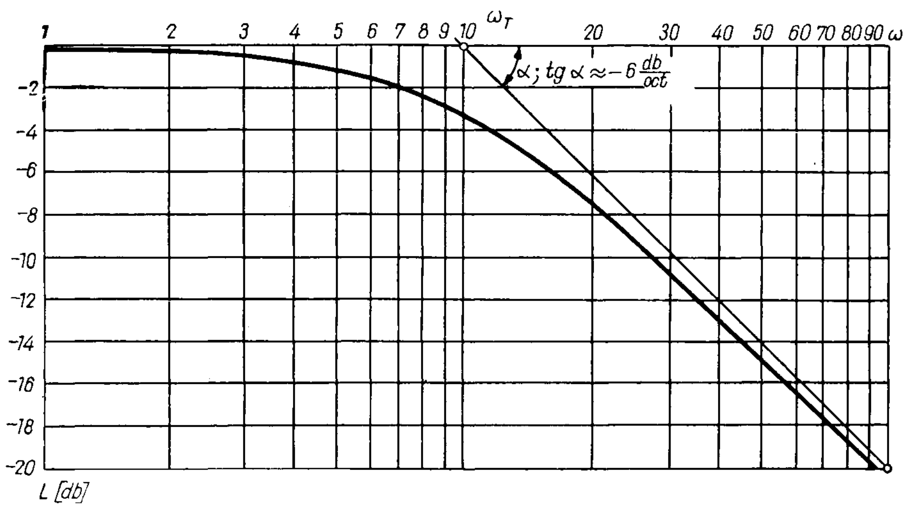


FIG. 98

the point $\omega = \omega_T$ with a slope of -6 dB per octave. Indeed

for $\omega = \omega_1$ $\qquad L = L_1 = 20 \log \omega_T - 20 \log \omega_1$,

and for $\omega = \omega_2 = 2\omega_1$

$$L = L_2 = 20 \log \omega_T - 20 \log \omega_1 - 20 \log 2 ,$$

i. e.

$$L_2 \approx L_1 - 6 .$$

We can usually substitute a broken line consisting of sections of the asymptotes for the true logarithmic amplitude characteristic of a single-capacitance stage since the curve differs little from this line (Fig. 98).

Let us estimate the amount of error obtained in this substitution:

The error is
$$\delta = L_{\text{true}} - L_{\text{app}}$$

where L_{true} is the true value of the characteristic, L_{app} is the approximate value obtained in replacing the characteristic by its asymptote.

For $\omega \ll \omega_T$
$$\delta = 20 \log \omega_T - 20 \log \sqrt{\omega_T^2 + \omega^2}.$$

For $\omega \gg \omega_T$
$$\delta = 20 \log \omega - 20 \log \sqrt{\omega_T^2 + \omega^2}.$$

The values of δ for various frequencies are set out in Table II.

TABLE II.

ω	$\frac{1}{5}\omega_T$	$\frac{1}{4}\omega_T$	$\frac{1}{2}\omega_T$	ω_T	$2\omega_T$	$4\omega_T$	$5\omega_T$
δ in dB	—0·17	—0·3	—1	—3	—1	—0·3	—0·17

For $\omega = \omega_T$ the error is at its greatest and reaches -3 dB; for $\omega = 2\omega_T$ the error decreases to 1 dB. Thus, the absolute values of the error are insignificant.

Oscillatory stage. The transfer function of this stage is of the form

$$w(p) = \frac{k}{T'^2 p^2 + T_k p + 1}.$$

We transform this expression, dividing numerator and denominator by T'^2:

$$w(p) = \frac{\frac{1}{T'^2} k}{p^2 + \frac{1}{T'^2} T_k p + \frac{1}{T'^2}}.$$

Let us put

$$\frac{1}{T'} = \omega_d, \quad \frac{T_k}{T'} = 2\xi \quad \text{or} \quad \xi = \frac{T_k}{2T'}.$$

Then

$$w(p) = \frac{k\omega_d^2}{p^2 + 2\xi\omega_d p + \omega^2}.$$

Putting $p = i\omega$, we obtain:

$$w(i\omega) = \frac{k\omega_d^2}{(\omega_d^2 - \omega^2) + 2i\xi\omega_d\,\omega},$$

which gives

$$A(\omega) = \frac{kw_d^2}{\sqrt{(\omega_d^2 - \omega^2)^2 + 4\xi^2\,\omega_d^2\,\omega^2}}. \tag{2.57}$$

Putting the value of $A(\omega)$ obtained from(2.57) in $L = 20 \log A(\omega)$ we obtain:

$$L = 20 \log \frac{k\omega_d^2}{\sqrt{(\omega_d^2 - \omega^2)^2 + 4\xi^2\,\omega_d^2\,\omega^2}}, \tag{2.58}$$

giving

$$L = 20 \log k + 20 \log \omega_d^2 - 20 \log \sqrt{(\omega_d^2 - \omega^2)^3 + 4\xi^2\,\omega_d^2\,\omega^2}.$$

For $k = 1$

$$L = 20 \log \omega_d^2 - 20 \log \sqrt{(\omega_d^2 - \omega^2)^2 + 4\xi^2\,\omega_d^2\,\omega^2}.$$

This is the exact equation of the logarithmic amplitude characteristic of an oscillatory stage. To simplify its construction in this case too we can replace it by its asymptotes.

Suppose that $\omega \to 0$. Then

$$L \to 20 \log \omega_d^2 - 20 \log \omega_d^2 \to 0.$$

Replacing the $\to$ by equals signs, we obtain the equation of the first asymptote.

Let $\omega \to \infty$. Then

$$\sqrt{\omega^4 + 4\xi^2\,\omega_d^2} \to \omega^2, \quad \text{since}\ \omega^4 \gg 4\xi^2\,\omega_d^2\,\omega^2$$

for sufficiently large values of ω.

Then

$$L \to 20 \log \omega_d^2 - 20 \log \omega^2$$

or, otherwise,

$$L \to 40 \log \omega_d - 40 \log \omega\ .$$

Replacing the arrows by equals signs we obtain the equation of the second asymptote, which itself represents a straight line passing

through the point corresponding to $\omega = \omega_d$. Let us therefore find its slope. If $\omega = \omega_1$, then

$$L_1 = 20 \log \omega_d^2 - 40 \log \omega_1 .$$

For
$$\omega = \omega_2 = 2\omega_1$$

$$L_2 = 20 \log \omega_d^2 - 40 \log \omega_1 - 40 \log 2 .$$

Taking into account that $20 \log 2 \approx 6$ we obtain

$$L_2 \approx L_1 - 12 .$$

Thus, the slope of the second asymptote is equal to 12 dB per octave.

Let us estimate the magnitude of the error $\delta = L_{\text{true}} - L_{\text{app}}$ which is introduced by replacing the characteristic of the oscillatory stage by its asymptotes, and let us find under what conditions this error will be negligible.

For $\omega < \omega_d$

$$\delta = 20 \log \omega_d^2 - 20 \log \sqrt{(\omega_d^2 - \omega^2)^2 + 4\xi^2 \omega_d^2 \omega^2} .$$

For $\omega > \omega_d$

$$\delta = - 20 \log \sqrt{\omega_d^2 - \omega^2)^2 + 4\xi^2 \omega_d^2 \omega^2} + 20 \log \omega^2 .$$

Obviously, the error depends on the quantity ξ. To estimate the errors we construct a family of curves for various values of the parameter ξ (Fig. 99).

For $\xi = 0{\cdot}6$ the error is little different from zero. It does not exceed roughly 3 dB for $0{\cdot}4 < \xi < 0{\cdot}7$.

If $\xi < 0{\cdot}4$, then near the value $\omega = \omega_d$ the amplitude characteristic must be calculated from the exact formula, since for small values the error may considerably exceed 3 db and may tend to infinity as $\xi \to 0$, i. e. when the oscillatory element approaches the conservative stage.

An open loop consisting of astatic, oscillatory and single-capacitance stages. For such a circuit the transfer function has the form

$$W(p) = \frac{k}{p} \prod_{j=l}^{m} \frac{k_j}{T_j p + 1} \prod_{j=l}^{n} \frac{k_j}{T_t'^2 p^2 + T_{kj} p + 1} .$$

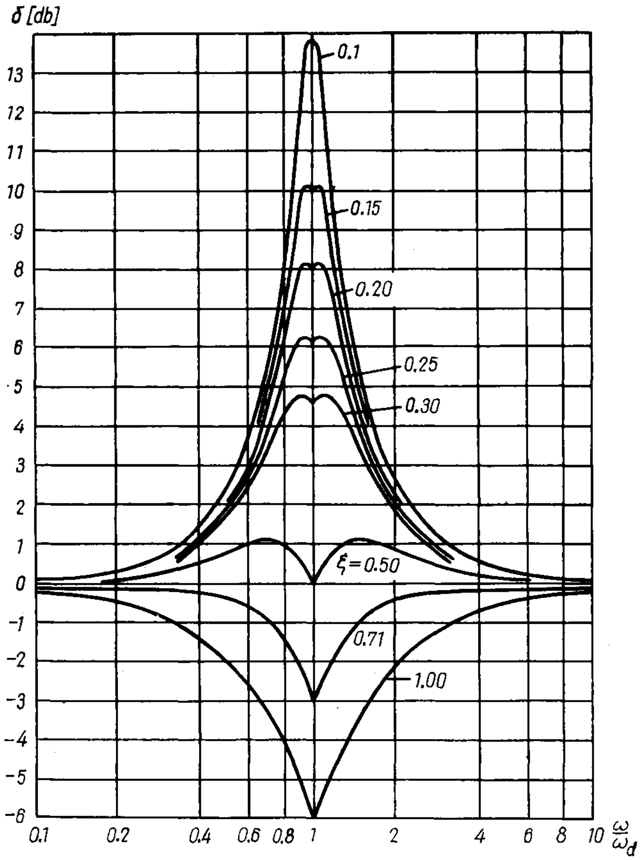


FIG. 99

The amplitude characteristic is

$$A(\omega) = |W(i\omega)| = \frac{k}{|i\omega|} \prod_{j=1}^{m} \frac{k_j}{|T_j i\omega + 1|} \prod_{j=l}^{n} \frac{k_j}{|(1 - T_j'^2 \omega^2) + iT_{kj}\omega|}$$

or

$$L = 20 \log A(\omega) = 20 \log \frac{k}{|i\omega|} \sum_{j=p}^{m} \log \frac{k_j}{|T_j i\omega + 1|} +$$

$$+ 20 \sum_{j=1}^{n} \log \frac{k_j}{|(1 - T_j'^2 \omega^2) + iT_{kj}\omega|} .$$

In order to obtain an approximate construction of the logarithmic amplitude characteristic of an open single-loop system, consisting of

single-capacitance, oscillatory and astatic stages, the following operations must be carried out.

1. As abscissa we take ω on the logarithmic scale, and as ordinate L in decibels.

2. Along the ω-axis we mark the points equal to $\omega_T = \frac{1}{T}$ and ω_d (conjugate to the stage frequencies). For an astatic stage, it is conventional to put $\omega_T = 1$.

3. Through each point corresponding to a conjugate frequency we draw a straight line with a slope of 6 dB per octave (if the stage is astatic of single-capacitance) or of 12 dB per octave (if the stage is oscillatory). We must then add the ordinates of all the broken lines obtained.

4. In order to make the curve more accurate we have to take into account the accumulated errors introduced by representing the curve of the element by its asymptotes.

For an astatic stage, the error is equal to zero. For a single-capacitance stage it is given in Table II, and for an oscillatory stage it is found graphically (see Fig. 99).

When, for the oscillatory stage $0{\cdot}4 < \xi < 0{\cdot}7$, the error can be taken roughly as not exceeding 3 dB. Usually sufficient accuracy is obtained if we regard the error for each conjugate frequency as being obtained only from the stage with this conjugate frequency. Then for each ω_T and ω_d we must plot a point 3 dB. below the break. The points thus obtained must be joined by a smooth curve, as shown in Fig. 100.

5. This procedure was done for $k = 1$. If $K = \Pi\, k_j \neq 1$, then the whole curve must be raised by the amount $L = 20 \log K$.

Stages with derivative action. We suppose that one of the stages of the considered circuit contains derivative action. Let the inherent operator of this stage be $d\,(p)$. Then the Laplace transformed differential equation of this stage can be written in the following form:

$$d\,(p)\,L\,[x_{\text{out}}] = k(1 + \varrho p)\,L\,[x_{\text{in}}]\,.$$

The usual rules give us the transfer function

$$w\,(p) = k\frac{1 + \varrho p}{d\,(p)}\,.$$

By definition

$$A(\omega) = |w(i\omega)| = k \frac{|1 + \varrho i\omega|}{|d(i\omega)|}$$

and hence

$$L = 20 \log A(\omega) = 20 \log k + 20 \log |1 + i\varrho\omega| - 20 \log |d(i\omega)|$$

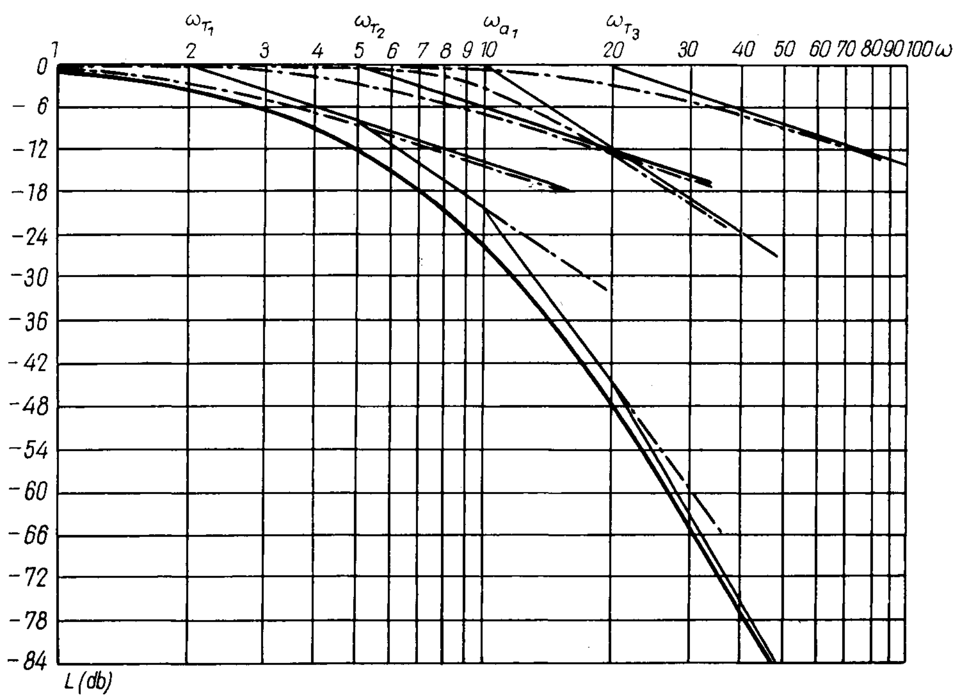


FIG. 100

From this expression we see that to construct the amplitude logarithmic characteristic of a stage with derivative action we must add the ordinates of the curve

$$20 \log |1 + i\varrho\omega|,$$

to the ordinates of the logarithmic amplitude characteristic of the stage without derivative action, i.e. we can first construct the characteristic of the stage with the transfer function $\frac{k}{d(p)}$, then construct

the characteristic of the equivalent stage with the transfer function $(1 + \varrho p)$, and then add them.

Let us construct the characteristic of an equivalent stage with the transfer function

$$w(p) = 1 + \varrho p \, . \tag{2.59}$$

We transform the expression (2.59) as follows

$$w\,(i\omega) = 1 + i\omega\varrho = \frac{i\omega + \omega_\varrho}{\omega_\varrho}\,, \quad \text{where } \omega_\varrho = \frac{1}{\varrho}\,.$$

But, by definition,

$$A\,(\omega) = |\,w\,(i\omega)\,| = \sqrt{\frac{\omega_\varrho^2 + \omega^2}{\omega_\varrho^2}}\,.$$

Hence

$$L = 20 \log A\,(\omega) = 20 \log \sqrt{\omega_\varrho^2 + \omega^2} - 20 \log \omega_\varrho .$$

In order to construct this curve we find its asymptotes as before. For small values of the frequencies, i.e. for $\omega \to 0$, $L \to 0$ also. Thus, the first asymptote will be the x-axis. For large values of frequency we obtain the equation of the asymptote in the following form

$$L = 20 \log \omega - 20 \log \omega_\varrho \, .$$

This expression is the equation of a straight line, but now with a positive slope. We can find the slope of this asymptote as before. It is equal to $+6$ dB per octave.

We note, furthermore, that for the second asymptote when $L = 0$ $\omega = \omega_\varrho$, i.e. the asymptote intersect at the point $\omega = \omega_\varrho$.

The errors introduced by substituting the points of the curve by those of its asymptotes can be determined of the table which was used for the single-capacitance stages, except that they will be of the opposite sign.

Figures 101 and 102 give examples of the construction of the logarithmic frequency characteristics of single-capacitance (Fig. 101) and oscillatory (Fig. 102) stages having first derivative action.

Let us suppose that in one of the circuit stages there is second derivative action. In this case, the transfer function has the form

$$w(p) = k\frac{w_1(p)}{w_2(p)} = k\,\frac{(1 + ep + sp^2)}{d(p)}$$

and

$$A(\omega) = |w(i\omega)| = k\left|\frac{(1 + i\varrho\omega - s\omega^2)}{d(i\omega)}\right|$$

or, in logarithmic coordinates

$$L = 20\log A(\omega) = 20\log k - 20\log|d(i\omega)| + 20\log|1 + \varrho\, i\,\omega - s\omega^2|$$

From the last expression it is clear that to construct the logarithmic characteristic of this stage we must add together the ordinates of two logarithmic characteristics: that of the stage with transfer function $\frac{k}{d(p)}$ and that of a single equivalent stage with transfer function

$$\omega_1(p) = (1 + \varrho p + sp^2)\,.$$

Let us construct the logarithmic characteristic of this equivalent stage. We transform its transfer function by putting

$$\xi = \frac{\varrho}{2\sqrt{s}}\,,$$

$$\omega_d = \frac{1}{\sqrt{s}}\,.$$

Then

$$w_1(p) = \frac{p^2 + 2\xi\,\omega_d\, p + \omega_d^2}{\omega_d^2}\,.$$

By definition

$$A_1(\omega) = |w_1(i\omega)| = \sqrt{\frac{(\omega_d^2 - \omega^2)^2 + 4\xi^2\,\omega_d^2\,\omega^2}{(\omega_d^2)^2}}$$

and

$$L = 20\lg A_1(\omega) = 20\lg\sqrt{(\omega_d^2 - \omega^2)^2 + 4\xi^2\,\omega_d^2\,\omega^2} - 20\log\omega_d^2\,.$$

We met similar equations when we considered the logarithmic characteristics of oscillatory stages. This equation differs from the earlier ones only in the sign of all its terms. We can therefore assert that the asymptotes of this curve will be the straight line $L = 0$ for small frequencies and the straight line

$$L = 20 \log \omega^2 - 20 \log \omega_d^2$$

for large frequency values.

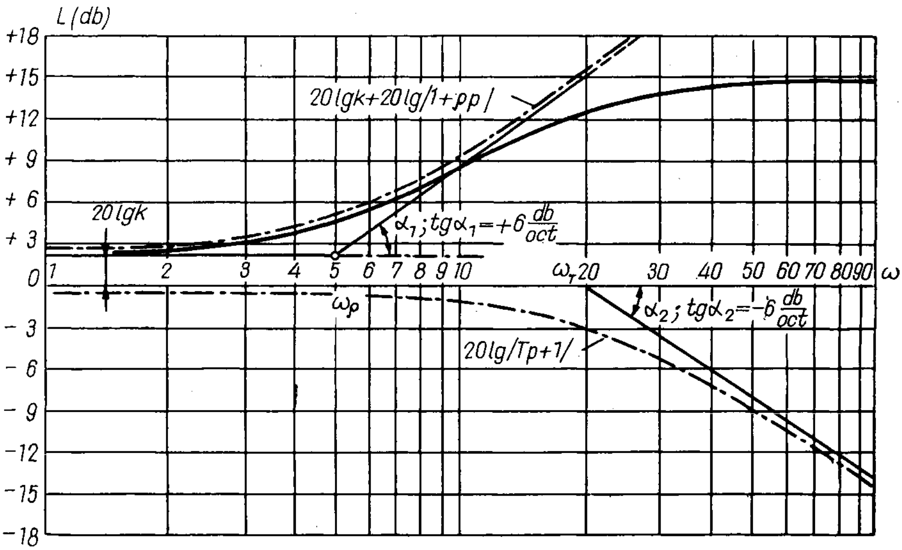


FIG. 101

The slope of the second asymptote will be positive and equal to 12 dB per octave. The errors introduced in the substitution of the exact curve by its asymptotes will be given by the same graphs as for the ordinary oscillatory stages, but their sign will be reversed.

Figure 103 is an example of the construction of the logarithmic characteristic of an oscillatory stage with first and second derivative action.

Single-loop circuit with derivative action. Logarithmic characteristics in this case are constructed according to the same rules as in the ordinary single-loop circuit. The only difference is that it is now necessary to take into account the logarithmic characteristic of the equivalent stage effecting the derivative action.

The construction of phase and amplitude-phase characteristics. It was shown above that the construction of the amplitude characteristic of a single-loop circuit is simplified when we use logarithmic coordinates, since the characteristics of the stages may be replaced by a broken straight line, and the characteristics of the open system in logarithmic coordinates are formed by the sum and not the product of the ordinates.

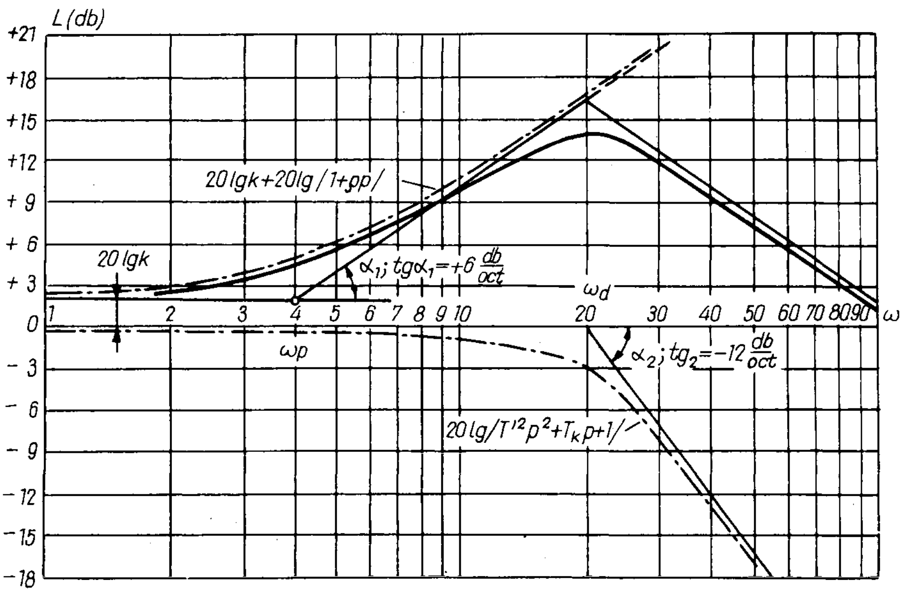


FIG. 102

The phase characteristic of a system in ordinary (non-logarithmic) coordinates also is obtained by summing the phase characteristic of its elements, since the argument of the product of vectors is equal to the sum of the arguments of the factors. Let

$$W(p) = \prod_{j=1}^{n} \frac{k_j(p)}{d_j(p)} .$$

Then

$$\arg W(i\omega) = \sum_{j=1}^{n} \arg k_j(i\omega) - \sum_{j=\lambda}^{n} \arg d_j(i\omega) .$$

Here all the $k_j(p)$ and $d_j(p)$ are polynomials of not higher than the second order. The values of their corresponding arguments are given in Table III.

TABLE III

$k(p)$ or $d(p)$	$\arg k(i\omega)$ or $\arg d(i\omega)$
c	0
$bp + c$	$\arctan \dfrac{b\omega}{c}$
$ap^2 + bp + c$	$\arctan \dfrac{b\omega}{c - a\omega^2}$

In Table III a, b and c are constants.

Thus, the construction of the phase characteristic of the separate stages of the open system also presents no difficulty.

Knowing the logarithmic amplitude characteristics and phase characteristics it is easy to construct from them the amplitude-phase characteristic of the system.

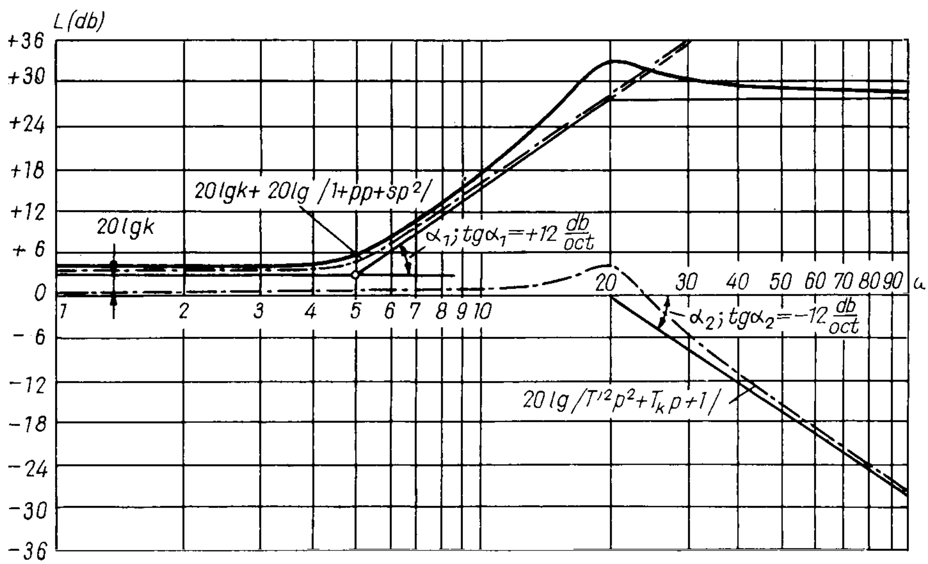


FIG. 103

(f) *General remarks about the use of the frequency characteristics of a system to determine its properties*

For non-linear elements (whether they are linearizable or not), the essential difference between experimentally determined frequency characteristics and frequency characteristics calculated from the equations for the same conditions, must be borne in mind.

When deriving the equation of motion we replaced the non-linear characteristics by their tangents at the point determined by the given conditions. When determining the frequency characteristics experimentally, we "average" the non-linear functions and replace the non-linear curves not by their tangents, but by secants, whose slope changes when there is a change in the amplitude of the sinusoidal disturbance applied at the input of the element.

Thus, only when we start from the equations of motion can a strict transition to a consideration of the system "for small oscillations" be effected, in the sense laid down in the theory of oscillations. When we are looking for the properties of a system with the use of frequency characteristics, the strict formulation of the problem of the investigation of the system "for small oscillations" is not infringed, only if they calculated from the equations of motion linearized by the method in this chapter. In practice, the time constant of the element or its coefficient of amplification is sometimes determined from a time characteristic which has been found by experiment. Of course, constants found in this way which enter into the equation of the linear model of the element also "average" the non-linear functions.

When determining the properties of a system both from experimental frequency characteristics and when the equations given have coefficients obtained by averaging the time characteristics, the problem is not correctly put: we consider the oscillations of the equivalent averaged linear model of the given system, and not small oscillations of the system in the exact sense of the term. Such a non-rigorous statement of the problem is justified only by the accumulation of experience which shows that often, in spite of it, the results obtained are admissible. Of course, this does not prove that correct results are obtained in all cases, and therefore in the experimental averaging of non-linearities (when finding the frequency characteristics or when determining the coefficients of the equations of motion from the time characteristics) great care and circumspection are required.

9. Concluding Remarks

Let us summarize the contents of this chapter.

The elements of automatic control systems are usually non-linear. In the linear theory of automatic control we consider not the real system, but its linear model. If the elements are linearizable, then a study of the linear model enables us to estimate the control process in the real system, but only for sufficiently small disturbances. The linear model depends not only on the properties of the system, but also on the conditions of operation: for each value of the load on the object, of the controller tuning, and so on, a separate linear model must be constructed and each possible set of operational conditions must be studied separately. We therefore begin the preparation of material for a dynamic calculation with the construction of the static characteristics of the system and the determination of the equilibrium values of the coordinates for all operational conditions.

The transfer function of the system or its frequency characteristic can be used as the initial material for a linear analysis.

In each case the calculation is begun by breaking the system down into its elements, choosing generalized coordinates and their origin and direction. A static investigation leading to the construction of the static characteristic of the system as a whole from the static characteristics of its elements, is made; this gives a family of curves determining the steady values of all the coordinates of the system for any possible set of conditions (i.e. for a series of fixed values of the load, tuning, etc.).

Three cases can arise:

(1) the properties of all the elements are determined by the equations of motion;

(2) the properties of all the elements are obtained from experimentally derived frequency characteristics under all given conditions;

(3) some of the elements have experimentally derived frequency characteristics, while the others do not.

In the first case the next step in preparing the material for the problem is the derivation and linearization of the initial equations of motion for all the system elements. The linearized equations of the elements are reduced to typical forms by changing to relative coordinates and by introducing time constants and amplification coefficients into the problem. Then, depending upon the action circuit of

the given system, the transfer function of the system for all possible conditions is derived. For different conditions, the transfer functions usually differ only in the numerical values of their coefficients, and they can be calculated for all the given conditions at once.

In the second case, when the frequency characteristics of each system element for each set of conditions are given, the frequency-characteristic of the system for any conditions is formed from them.

In the third case, when only some of the elements have experimentally determined frequency characteristics, the frequency characteristic of the system as a whole is also used as initial material for the linear calculations. In order to construct it, we must form the equations of motion for those elements without frequency characteristics, we must linearize them and reduce the linearized equations to typical form. We then form the transfer function of each element and from this construct its frequency characteristic. In some cases, the construction of the frequency characteristics can be made easier by using logarithmic characteristics.

Then, just as in the second case, the frequency characteristic of the system can be constructed from the frequency characteristics of all the stages. Here we must construct as many characteristics as there are possible sets of conditions.

Sometimes, even when there are no experimental frequency characteristics at all, the result of the preliminary work for the calculation is presented in the form of frequency characteristics of the system under all conditions. This happens in those cases when it is wished to carry the linear analysis further by frequency methods. In such cases, after the transfer functions of the system have been obtained for various sets of conditions, the frequency characteristics are formed from them separately for each set of conditions.

CHAPTER III

THE STABILITY OF THE LINEAR MODEL OF AN AUTOMATIC CONTROL SYSTEM

IN THE majority of cases the problem of control consists of establishing and maintaining over a period of time the operating state of the controlled object. This problem gives rise to the requirement that the system of automatic control should possess a definite stability, even if only when the disturbances which affect the operating state of the controlled object or which act on the elements of the controller, are small.

Let us suppose that some external action has disturbed equilibrium in the system and has set the control process into motion. Further, suppose that this action then disappears. Then the control process is called *stable* or *of steady oscillation* if, after a sufficiently small disturbance, the action of the controller results in the restoration of the same state as that maintained by the controller up to the time of the disturbance. In the contrary case, the control process is said to be *unstable* or *of increased oscillation.**

According to this definition, the stability of the process depends on the action of the controller after the disturbance has ceased to act on the system. At that moment, the system deviated in a definite way from its equilibrium state, and these deviations may be taken as the initial deviations. Then the concept of a "stable" or "unstable" control process may be formulated in the following way: a process is said to be stable if, for any sufficiently small initial deviations, the equilibrium is restored to the control system as a result of the action of the controller; the process is said to be unstable if we can find deviations as small as we please for which the controller does not

* The concept of "stable process" will be made more precise in Section 4 of this chapter.

restore operating state which existed in the system before the appearance of these initial deviations.

We will show further that for the linear model of a system of automatic control, the fact that a small initial deviation figures in the above formulation of the concept of stability is of no importance: if the linear model is stable with respect to small disturbances then it is stable also with respect to any other disturbance, not necessarily small. But if we use this linear model to estimate the behaviour of the true system, then this is only valid with respect to small disturbances and therefore, in the best case, stability of the given control system in the meaning of this term given above.*

In practice automatic control is frequently employed in unstable processes. In most cases instability is the scourge of control. For this reason the calculation of an automatic control system is begun with the elucidation of its stability conditions.

1. Estimating the Stability of the Linear Model of a System, Using its Transfer Function

(a) *General considerations*

Let us suppose that at time $t = 0$ the control system we are considering is in the position of equilibrium, that the control process is called into action because a certain action is applied to one of the elements of the system and that at time $t = t_1$ this action disappears.

Let us take $t = t_1$ as the time coordinate origin. Then any coordinate of the system changes for $t > t_1$ according to the rule,**

$$x = \sum_{k=1}^{n} C_k \, e^{p_k t} , \tag{3.1}$$

where the C_k are constants, defined in the way described in Appendix I, and p_k are the roots of the characteristic equation of the system which are obtained as in the previous chapter by equating the denominator of the transfer function of the system to zero.

* At the end of this chapter (in Section 4) we shall investigate in detail the connexion between the stability of the original system and that of its linear model. The definition of the term "stability" given here differs from the classical definition of Lyapunov and corresponds to the concept "steady oscillation process" often used in technology.

** See Appendix I

If all the roots p_k of the characteristic equation are real numbers, then the C_k also are all real numbers, and x is the sum of exponential functions. If all the p_k are negative, then all the exponential functions, and hence x also, tend to zero as $t \to \infty$; however, if any one of the p_k is positive, then $x \to \infty$.

Now let the characteristic equation have complex-conjugate roots p_k and p_{k+1}. Then each sum $C_k e^{p_k t} + C_{k+1} e^{p_{k+1} t}$ on the right-hand side of (3.1) may be replaced (by Euler's identity) by the sum

$$e^{\alpha_k t} (A_k \sin \beta_k t + B_k \cos \beta_k t), \tag{3.2}$$

where α_k is the real part and β_k the imaginary part of the complex root $p_k = \alpha_k + i \beta_k$; and A_k and B_k are constants.

Indeed, from Euler's identity

$$e^{iz} = \cos z + i \sin z$$

we have

$$C_k e^{(\alpha_k + i\beta_k)t} = C_k e^{\alpha_k t} (\cos \beta_k t + i \sin \beta_k t),$$

$$C_{k+1} e^{(\alpha_k - i\beta_k)t} = C_{k+1} e^{\alpha_k t} (\cos \beta_k t - i \sin \beta_k t).$$

Hence

$$C_k e^{p_k t} + C_{k+1} e^{p_{k+1} t} =$$
$$= e^{\alpha_k t} [(C_k + C_{k+1}) \cos \beta_k t + i (C_k - C_{k+1}) \sin \beta_k t].$$

We note that C_k and C_{k+1} are complex-conjugate numbers. Let $C_k = \psi + i \varphi$. Then $C_{k+1} = \psi - i \varphi$. Therefore $C_k + C_{k+1} = 2\psi$ and $i(C_k - C_{k+1}) = -2\varphi$ are real. Denoting them by A_k and B_k we obtain (3.2).

The amplitudes $A_k e^{\alpha_k t}$ of the terms of (3.2) entering into (3.1) tend to zero if all the real parts α_k of the complex roots of the characteristic equation are negative. The solution of (3.1) grows without limit if the real part α_k of even one root is positive.

The zero value of the coordinates is their value in the position of controlled equilibrium. If therefore a coordinate tends to zero as $t \to \infty$, this means that equilibrium is restored.

From this it follows that *the necessary and sufficient condition for the stability of the control process in its linear approximation (i.e. in the case when the process is described by a system of linear equations) is that all real roots of the characteristic equation be negative and all complex roots have a negative real part.*

In the plane of the complex variable p let us draw the points corresponding to all the roots of the characteristic equation (Fig. 104). The system is stable if all n points lie to the left of the imaginary axis, and is unstable if any one point lies to the right of the imaginary axis. We agree to say that the system is on the border of stability when any one of the points lie on the imaginary axis (i.e. if the characteristic equation has any zero root or any one pair of purely imaginary roots).

Thus, in order to estimate the stability, it is not necessary to calculate the roots of the characteristic equation. It is necessary only to discover whether they all lie to the left of the imaginary axis or not.

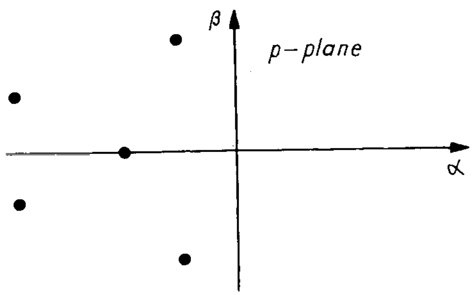


FIG. 104

Usually two statements of this problem are found. Firstly, we may be given certain parameters, and have to find for which values of the remaining parameters the system is stable. Secondly, we may consider all the parameters to be given, and then determine whether or not the system is stable for these parameter values.

The first problem is solved by constructing the region of stability, and the second by using the stability criteria.

(b) *The general concept of the D-partition*

We consider the characteristic equation

$$a_0 p^n + a_1 p^{n-1} + \ldots + a_{n-1} p + a_n = 0.$$

The aggregate of values $a_0, a_1, a_2 \ldots, a_n$ may be interpreted geometrically as a point in n-dimensional space, with axes for the values of the coefficients $a_0, a_1, \ldots, a_n$.

To each point of this space there corresponds a definite value of the coefficients a_0, a_1, ..., a_n and consequently, definite values of all the roots p_1, p_2, ... , p_n of the characteristic equation.

If in this space there exists a region in which each point corresponds to a characteristic equation, all of whose roots lie to the left of the imaginary axis in the root plane, then the hypersurface bounding this region is called the *boundary of the region of stability.*

When there are only two coefficients, this region is bounded by a plane, when there are three by a three-dimensional surface and so forth.

Since all the coefficients a_k are determined by the values of the parameters of the differential equations of the system (time constants and amplification coefficients) we may clearly similarly construct the space of the parameters T_1, T_2, ... , k_1, k_2, ... and isolate a region of stability within it.

In practice we try to make it sufficient to construct the region of stability on a numerical straight line (one parameter) or in a plane (two parameters).

Let us consider by way of example a characteristic equation in which all the coefficients apart from two (say a_0 and a_n) are given.

Let us suppose that for some definite values of a_0 and a_n in the plane of the roots (in the p-plane) the given equation has k roots lying to the left and $n-k$ to the right of the imaginary axis (Fig. 105a). It is obvious that there always exists a curve (in general a hypersurface) which, in the plane of a_0 and a_n (in general case in the coefficient space) bounds a region (Fig. 105b) in which each point defines a polynomial also having k roots lying to the left and $n-k$ roots to the right of the imaginary axis. We denote the region bounded by this curve (in the general case, surface) by D $(k, n - k)$. Here k can be any whole number between 0 and n, and in this way, in the plane of a_0 and a_n (in the general case in the coefficient space) we can find regions D $(k, n - k)$ corresponding to various values of k. Thus, for example, if the characteristic equation is of the third degree, i. e. if $n = 3$, then in the general case the regions D (0. 3), D (1. 2), D (2. 1) and D (3. 0) can be found in the coefficient space. The last region is in fact also the region of stability in the coefficient space.

If in this characteristic equation all the coefficients except two are specific numbers, then the plane of these two undetermined coefficients may not contain any of the regions mentioned. If, for ex-

ample, there exists no region D (3. 0) then this means that for any values of these undetermined coefficients and for the given values of the remaining coefficients, the equation cannot have three roots with a negative real part (to the left of the imaginary axis).

The partition of the coefficient space of the characteristic equation into regions corresponding to the same number of roots lying to the left of the imaginary axis is called the *D-partition.* We note that the whole D-partition is realized by one surface (or line in the case of a two-dimensional coefficient space), this surface (line) dividing up

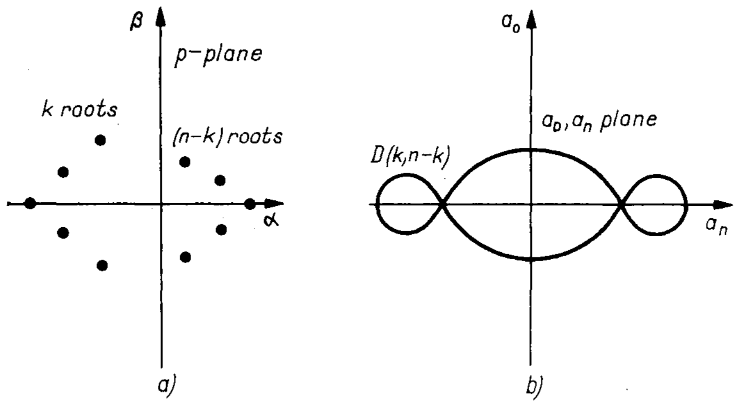


FIG. 105

all regions corresponding to the polynomials having k roots to the left of the imaginary axis for any number k.

Let us suppose that k roots of the polynomial

$$a_0 p^n + a_1 p^{n-1} + \ldots + a_{n-1} p + a_n = 0$$

lie to the left of the imaginary axis. Let us vary the values of the coefficients a continuously. The roots may then cross over to the right-hand half-plane, although clearly only by crossing the imaginary axis.* Consequently the imaginary axis in the p-plane is the reflection of the boundary of the D-partition, and the crossing of the latter in the coefficient space is represented by the roots in the root plane

* A root may also cross from the left-hand to the right-hand half-plane (i.e. the sign of its real part may change), when the real part goes off to infinity, i.e. when $a_0 \to 0$. This case will be considered separately below.

crossing the imaginary axis. This suggests the method for determining the D-partition boundary: its equation is found in parametric form by replacing p by $i\omega$ in the given polynomial (where ω is a variable). From this equation the boundary may be constructed if the value of ω is varied from $-\infty$ to $+\infty$.

We considered above the D-partition in the space of the coefficients of the characteristic equation. Of course, in a similar way, we can construct the D-partition for the space of any parameters on which the coefficients of the characteristic equation depend (for example, for the space of the time constants and amplification coefficients).

The aim of the analysis is to isolate the region of stability, the region $D(n, 0)$, but the construction of all the D-partitions often proves to be a very simple method for determining the boundary of the region of stability.

(c) *The construction of the stability region in the plane of one complex parameter*

Let us denote by λ the parameter whose value is varied in order to guarantee stability. Let us suppose that we may solve the characteristic equation with respect to λ, i.e. reduce it to the form

$$Q(p) + \lambda R(p) = 0 \quad \text{or} \quad \lambda = -\frac{Q(p)}{R(p)}.$$

Thus, for example, in the case of the equation

$$p^2 + a_1 p + a_2 = 0 \quad \text{and} \quad \lambda = a_1$$

we obtain

$$Q(p) = p^2 + a_2, \qquad R(p) = p.$$

In the case of the equation

$$(T_1 p + 1)(T_2 p + 1) + 1 = 0 \quad \text{and} \quad \lambda = T_1$$

we obtain

$$Q(p) = (T_2 p + 1) + 1, \qquad R(p) = p(T_2 p + 1).$$

Only real values of λ have any practical value. Let us suppose, however, for the time being, that λ is complex, and let us transform

the imaginary axis in the plane of the roots (the p-plane) into the λ-plane (Fig. 106). To do this we put $p = i\,\omega$ in the equation

$$\lambda = -\frac{Q(p)}{R(p)},$$

obtaining

$$\lambda(i\omega) = -\frac{Q(i\omega)}{R(i\omega)}$$

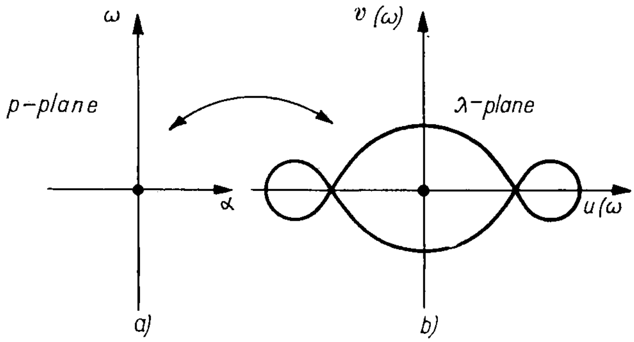


FIG. 106

and we separate real and imaginary parts

$$\lambda = -\frac{Q(i\omega)}{R(i\omega)} = u(\omega) + iv(\omega)\,.$$

Giving ω values from $-\infty$ to $+\infty$ we construct from these points a curve which is the transformation of the imaginary axis of the p-plane on λ-plane, i.e. the boundary of the D-partition in the λ-plane. This boundary in the given case is symmetrical with respect to the real axis, and in order to construct the whole curve it is sufficient to construct that half corresponding to $0 \leqslant \omega \leqslant +\infty$ and then to add to it its mirror reflection in the real axis.

If in the root plane we move along the imaginary axis from $-\infty$ to $+\infty$ (Fig. 107a) then the region where, for stability of the process, the roots must be distributed, will always be on the left.

Proceeding along the boundary curve of the D-partition from the point corresponding to $\omega = -\infty$, to the point corresponding to $\omega = +\infty$, we will shade this curve on the left (Fig. 107b). When in

the λ-plane the boundary of the D-partition is crossed from the shaded to the unshaded side, in the root-plane one root crosses the imaginary axis, passing from the left-hand half-plane to the right-hand half-plane.

Since only real values of λ are of practical use, of the whole region only the partition of the real axis is important.

Thus, in order to construct the boundary of the D-partition for any one parameter we must:

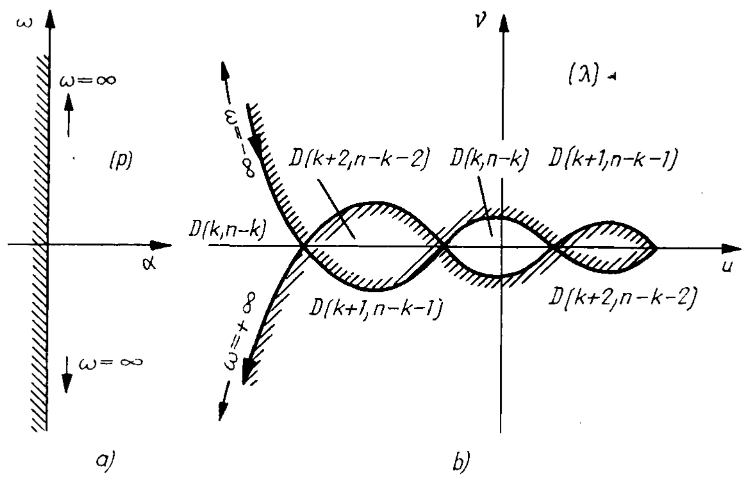


FIG. 107

(1) Solve the characteristic equation with respect to this parameter, i.e. reduce it to the form $\lambda = -\dfrac{Q(p)}{R(p)}$.

(2) Perform the substitution $p = i\omega$ and separate real and imaginary parts in the expression obtained, i. e. reduce it to the form

$$\lambda = u(\omega) + iv(\omega) .$$

(3) Setting out along the u and v axis their corresponding values, construct the curve obtained if ω is taken for all values from 0 to $+\infty$.

(4) Add to this curve its mirror reflection in the u-axis, i.e. its branch corresponding to $-\infty < \omega \leqslant 0$.

(5) Moving along this curve from the point corresponding to $\omega = -\infty$ to the point $\omega = +\infty$ shade the left-hand side of the curve.

If to any point of the λ-plane there corresponds a polynomial having k roots to the left of the imaginary axis, then the polynomial corresponding to any point to which we may pass over without crossing the D-partition boundary has the same number of roots to the left of the imaginary axis. If in passing to another point the curve is crossed from the non-shaded side to the shaded side, then this new point represents a polynomial having $k + 1$ roots to the left of the imaginary axis if the shading is single, and $k + 2$ roots if the shading is double (points of self-intersection of the curve).

If a polynomial with k roots to the left of the imaginary axis is represented by the point A and if from A we can reach the point B by crossing the boundary of the D-partition z_1 times from the shaded side and z_2 times from the unshaded side,* then B represents a polynomial having $(k + z_2 - z_1)$ roots to the left of the imaginary axis.

Thus, it is sufficient to know the distribution of the roots relative to the imaginary axis for any one arbitrary value of λ in order to determine the distribution for any other value of λ.

In practice, since we are only interested in real values of λ, having constructed the D-partition of the λ-plane and having determined the number of roots corresponding to each region, we have to determine which segment of the real axis belongs to the region of stability.

We give some examples of constructing the D-partition.

Example 1. The characteristic equation is given as

$$p^3 + p^2 + p + \lambda = 0.$$

Solving it for the parameter λ, we obtain

$$\lambda = -p^3 - p^2 - p.$$

Carrying out the substitution $p = i\,\omega$, we find

$$\lambda = i\,\omega^3 + \omega^2 - i\,\omega = u + iv,$$

where

$$u = \omega^2; \quad v = \omega^3 - \omega.$$

We construct the boundary of the D-partition in the λ-plane (Fig. 108.)

Passing along the curve from the point $\omega = -\infty$ to $\omega = +\infty$ we shade it on the left.

* Crossing across a point of self-intersection of the curve is here regarded as two intersections.

It is obvious that the region corresponding to the polynomials which have the greatest number of roots to the left of the imaginary axis will be the region containing the points of the real axis satisfying the inequality $0 < \lambda < 1$.

In order to verify that this region will be the region of stability, we consider the border point $\lambda = 0$.

For $\lambda = 0$

$$p_1 = 0;$$

$$p_{2,3} = -\frac{1}{2} \pm \sqrt{-\frac{3}{4}},$$

i.e. one root is equal to zero and two lie to the left of the imaginary axis.

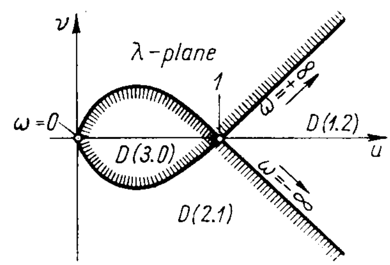


FIG. 108

Inside the given region the number of roots lying to the left of the imaginary axis must be greater by one, since in order to be in this region it is necessary to cross the boundary of the D-partition from the unshaded side. Hence polynomials having all three roots to the left of the imaginary axis correspond to this region.

Only real values of λ interest us and these are determined at once by the segment of the u-axis lying inside the region $D(3, 0)$. Thus, the considered system is stable only if $0 < \lambda < 1$.

EXAMPLE 2. Let us consider the characteristic equation

$$p^3 + \lambda p^2 + p + 1 = 0.$$

Solving it for λ, we find

$$\lambda = \frac{-p^3 - p - 1}{p^2}.$$

Putting $p = i\omega$ we obtain

$$\lambda = \frac{i\omega^3 - i\omega - 1}{-\omega^2} = \frac{1}{\omega^2} + i\left(\frac{1}{\omega} - \omega\right).$$

Giving ω values from $-\infty$ to $+\infty$, we construct the D-partition in the λ-plane (Fig. 109) and find the region which corresponds to the polynomials with the greatest number of roots lying to the left of the imaginary axis. In Fig. 109 this region is denoted by the letter R. At no point of the plane not

belonging to R can there be the same or a greater number of roots lying to the left of the imaginary axis*.

Let us find the number of roots to the left of the imaginary axis for $\lambda = 1$. In this case the characteristic equation reduces to the form

$$p^3 + p^2 + p + 1 = (p^2 + 1)(p + 1) = 0$$

and has roots

$$p_1 = -1; \; p_{2,3} = \pm i.$$

Thus for the point $\lambda = 1$ one root lies to the left of the imaginary axis, and two roots lie on it. Passing from the point $\lambda = 1$ to any point of the region R, we leave the curve on the shaded side, while at the point $\lambda = 1$, the shading

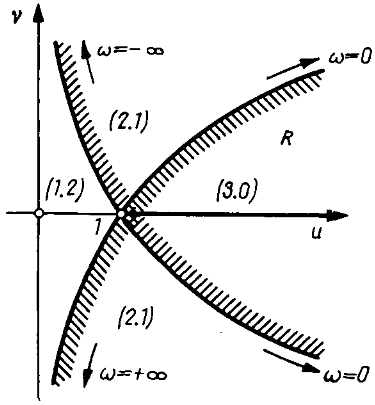


FIG. 109

is double (two branches of the D-partition boundary pass through this point). Consequently, the region R corresponds to the case when all three roots have a negative real part, i.e. it is the region of stability.

EXAMPLE 3. Let us determine what the time constant of viscous friction T_k of a controller must be if the motor having positive inherent regulation, time constant $T_1 = 1$ sec and coefficient of amplification $k_1 = 2$ is controlled by a static direct action controller with time constant $T_2' = 0{\cdot}1$ sec and amplification coefficient $k_2 = 20$.

The equations of motion for the motor and controller are

$$T_1 \frac{\mathrm{d}x_1}{\mathrm{d}t} + x_1 = -k_1 x_2; \quad T_2'^2 \frac{\mathrm{d}^2 x_2}{\mathrm{d}t^2} + T_k \frac{\mathrm{d}x_2}{\mathrm{d}t} + x_2 = k_2 x_1. \tag{3.3}$$

The characteristic equation has the form**

$$(T_1 p + 1)(T_2'^2 p^2 + T_k p + 1) + k_1 k_2 = 0.$$

* Indeed it is not possible to cross from any point of this region to any point not belonging to it without crossing the boundary of the D-partition from the shaded side to the unshaded side.

** See Chapter II.

After substituting the given values of T_1, T_2', k_1 and k_2 we obtain:

$$(p+1)(0{\cdot}01\,p^2 + T_k\,p + 1) + 40 = 0.$$

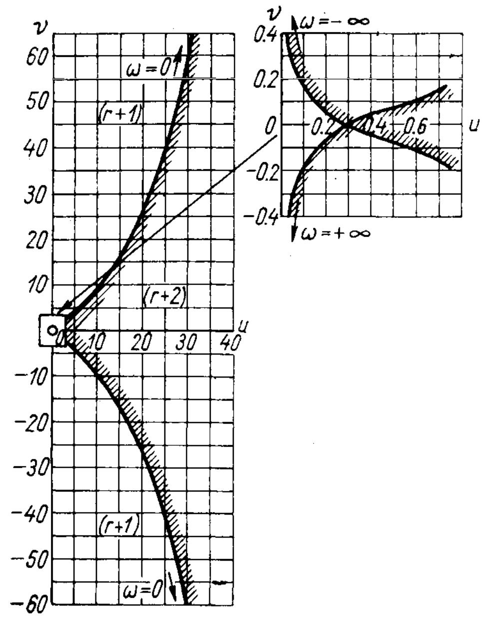


FIG. 110

Solving the last equation with respect to T_k, we find

$$T_k = -\frac{0{\cdot}01\,p^3 + 0{\cdot}01\,p^2 + p + 41}{p^2 + p}.$$

Substituting $i\omega$ for p, multiplying numerator and denominator by $(-i\omega - \omega^2)$, opening the brackets in the numerator and denominator, and collecting like terms we find

$$T_k = -\frac{0{\cdot}01\,i\omega^5 - 0{\cdot}99\,i\omega^3 - 41\,i\omega - 40\,\omega^2}{\omega^2 + \omega^4}$$

The construction of the D-curve from these points is shown in Fig. 110.

A system of direct control can always be made stable by the choice of T_k, and hence the T_k-plane must necessarily contain a stable region.

Let us suppose that for $T_k = 0{\cdot}1$ the equation (3.3) has r roots with a negative real part. Then the number of such roots for any other value of T_k (Fig. 110) is determined at once. Since we know beforehand that there exists a region of stability in the T_k-plane, then it can only be a region having the greatest mark, i. e. the region with a mark $r + 2$.

Thus, the considered system is stable for any $T_k \gtrsim 0{\cdot}3$.

(d) *The construction of the region of stability in the plane of two real parameters (the Vishnegradskii diagram)*

A Vishnegradskii diagram is taken to mean the plane of any two real parameters of a system in which the lines separating the region of stability are plotted. The Vishnegradskii diagram may thus be obtained by constructing the D-partition of the plane of the two real parameters, i.e. the plane section of the D-partition of the parameter space.

Let us suppose that the coefficients of the characteristic equation of the system

$$a_0 p^n + a_1 p^{n-1} + a_2 p^{n-2} + \dots + a_{n-1} p + n_n = 0$$

depend on two parameters* μ nad η, and let us restrict ourselves to the case when these parameters enter into the equation linearly, so that this equation can be reduced to the form

$$\mu S(p) + \eta Q(p) + R(p) = 0 .$$

For example, the equation

$$(\mu p + 1)(5p + 1) + 3\eta = 0$$

can be reduced to the form

$$\mu(5p^2 + p) + 3\eta + (5p + 1) = 0 .$$

In this case

$$S(p) = 5p^3 + p$$
$$Q(p) = 3$$
$$R(p) = 5p + 1$$

Putting, further, $p = i\omega$ and separating real and imaginary parts, we obtain

$$\mu S(i\omega) + \eta Q(i\omega) + R(i\omega) = u(\omega) + iv(\omega)$$

In the general case both functions $u(\omega)$ and $v(\omega)$ depend not only on ω, but also on the two parameters μ and η. In order to construct the boundary of the D-partition it is necessary to determine μ and η

* These parameters may be, in particular, simply two coefficients of the considered equation.

for each ω, by solving simultaneously the two equations

$$u(\omega) = 0\,,$$

$$v(\omega) = 0\,.$$

If in each of them we separate the terms containing μ and η, then we obtain a system of two equations with two unknowns:

$$u(\omega) = \mu\, S_1(\omega) + \eta\, Q_1(\omega) + R_1(\omega) = 0\,,$$

$$v(\omega) = \mu\, S_2(\omega) + \eta\, Q_2(\omega) + R_2(\omega) = 0\,.$$

Solving this system of two linear algebratic equations with respect to μ and η for each value to ω we obtain:

$$\mu = \frac{\begin{vmatrix} -R_1 & Q_1 \\ -R_2 & Q_2 \end{vmatrix}}{\begin{vmatrix} S_1 & Q_1 \\ S_2 & Q_2 \end{vmatrix}} = \frac{-R_1 Q_2 + R_2 Q_1}{S_1 Q_2 - S_2 Q_1}\,;$$

$$\eta = \frac{\begin{vmatrix} S_1 & -R_1 \\ S_2 & -R_2 \end{vmatrix}}{\begin{vmatrix} S_1 & Q_1 \\ S_2 & Q_2 \end{vmatrix}} = \frac{-S_1 R_2 + S_2 R_1}{S_1 Q_2 - S_2 Q_1}\,.$$

The equations $u(\omega) = 0$ and $v(\omega) = 0$ determine one value of μ and one value of η for each ω only when these equations are simultaneous and independent. If for some value of ω the numerator and denominator become zero, then for this value of ω one of the equations $u(\omega) = 0$ or $v(\omega) = 0$ is a consequence of the other, and for this value of ω we obtain not a point but a straight line in the μ, η-plane. In this case, either of the equations $u(\omega) = 0$ or $v(\omega) = 0$ is the equation of this straight line when this value of ω is substituted.

If the coefficient of the highest term of the characteristic equation depends on the parameters μ and η, then, by equating this coefficient to zero we obtain the equation of another straight line corresponding to $\omega = \infty$. These straight lines are called *singular*.

In order to shade the boundary of the D-partition we must move along the boundary in the direction of ω increasing, and shade it on the left edge at those points for which $\Delta > 0$ and on the right edge for those points for which $\Delta < 0$ where $\Delta = \begin{vmatrix} S_1 & Q_1 \\ S_2 & Q_2 \end{vmatrix}$. Usually the curve

is traversed twice: once when ω goes from $-\infty$ to 0, and once when it changes from 0 to ∞, but it is shaded both times on the same side, since usually the sign of Δ changes for $\omega = 0$ or $\omega = \infty$.* Singular lines pass through these points most frequently. They are shaded in this case as is shown in Fig. 111. Near the point of intersection of the curve and the straight line their shaded sides must be directed towards one another.

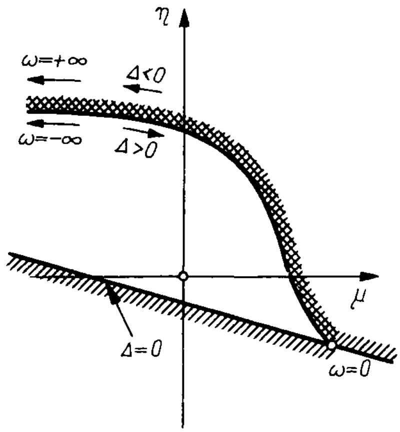


FIG. 111

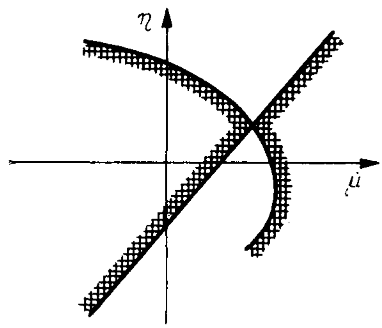


FIG. 112

There are rare exceptions when these singular lines must not be shaded (see Example 1). If Δ becomes zero (which is rare) for $\omega \neq \infty$ or $\omega \neq 0$ and the sign of Δ does not change, then these straight lines are not shaded and can generally be removed from the construction. If the sign of Δ does change, then the lines are shaded as in Fig. 112.

Finally, if $\Delta \equiv 0$, then the singular lines alone are the boundary of the D-partition.

Figure 113 gives examples of the shading of the D-partition curve and of the singular lines.

EXAMPLE 1. Given the characteristic equation

$$p^3 + \mu p^2 + \eta p + 1 = 0.$$

* In constructing the boundary of the D-partition the following rules must be observed:

(a) first write down the equation $u(\omega) = 0$, and then $v(\omega) = 0$;

(b) if μ is the first variable in these equations and η the second, then in constructing the curve the system of coordinates (μ, η) must be right-handed

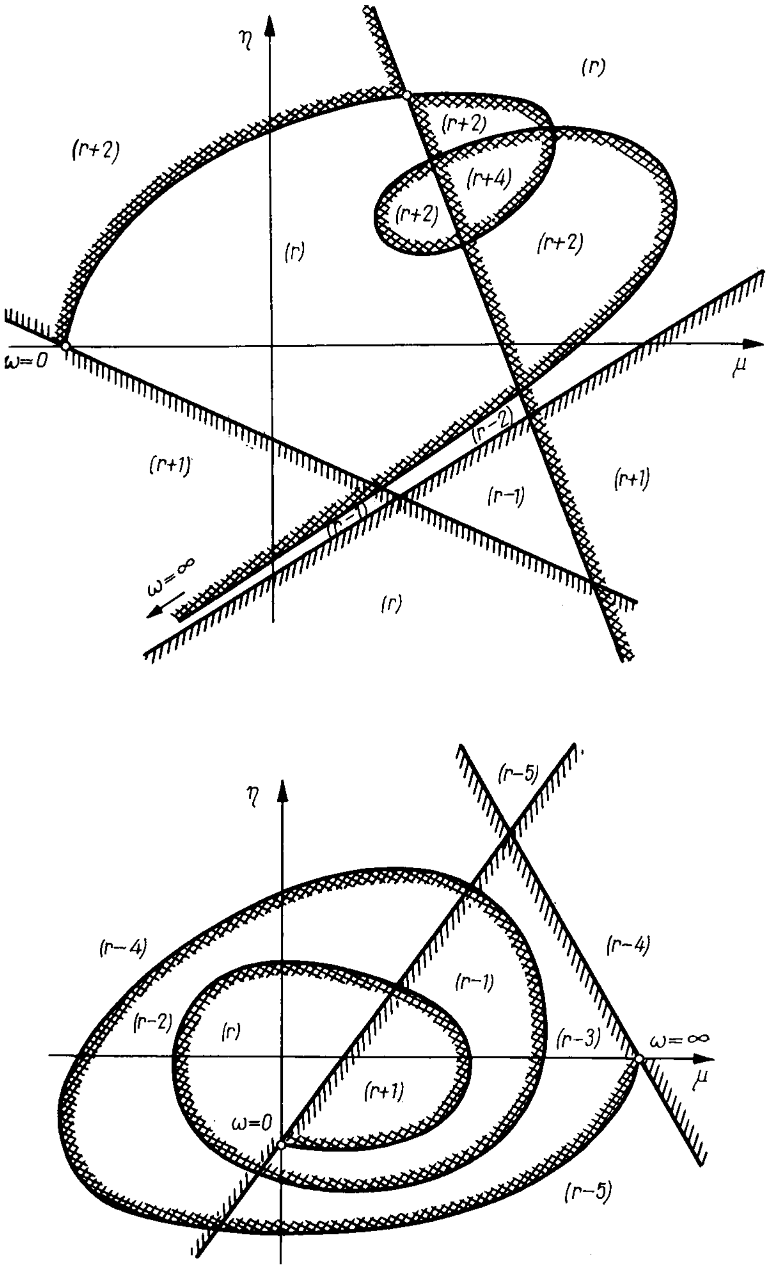


FIG. 113

We construct the D-partition by writing it in the form

$$\mu p^2 + \eta p + (p^3 + 1) = 0$$

and putting

$$p = i\omega .$$

Then

$$-\mu \omega^2 + \eta i \omega + 1 - i \omega^3 = 0$$

or

$$-\mu \omega^2 + 1 + i(\eta\omega - \omega^3) = 0 ,$$

whence

$$u(\omega) = \mu(-\omega^2) + \eta \cdot 0 + 1 = 0 ,$$

$$v(\omega) = \mu \cdot 0 + \eta\omega - \omega^3 = 0 .$$

The determinant of this system

$$\begin{vmatrix} -\omega^2 & 0 \\ 0 & \omega \end{vmatrix} = -\omega^3$$

is equal to zero only when $\omega = 0$.

For any $\omega \neq 0$

$$\mu = \frac{1}{\omega^2} \quad \text{and} \quad \eta = \omega^2 ,$$

therefore

$$\mu = \frac{1}{\eta} .$$

The boundary of the D-partition is a rectangular hyperbola (Fig. 114).

The point $\omega = 0$ lies at infinity on the μ-axis, and the points $\omega = +\infty$ and $-\infty$ lie at infinity on the η-axis. The determinant changes sign at the point $\omega = 0$. The hyperbola therefore has double shading, shown in Fig. 114.

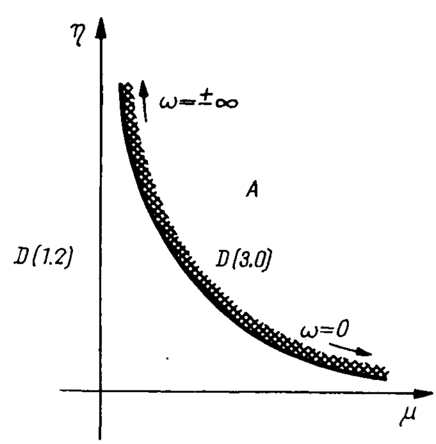


FIG. 114

The value $\omega = 0$ determines two singular straight lines: $\mu = \infty$ and $\eta = 0$. Only the line $\eta = 0$ will be of value, since the line $\mu = \infty$ does not divide a region in the finite part of the plane.

In the given case we need not shade the line $\eta = 0$. This is clear from the fact that to a point of this line $\mu = \eta = 0$, there corresponds an equation having one root to the left of the imaginary axis and two roots to the right, and no zero root. When the boundary of the D-partition with double shading is crossed, this corresponds to a change of sign in the real parts of the two roots. Hence the region A corresponds to all three roots lying to the left of the imaginary axis, i.e. is the region of stability.

EXAMPLE 2. Let us find what parameters (time constants and coefficients of amplification) of single-capacitance objects can be stably controlled by a static direct action controller having a time constant equal to 0·45 sec, a damping time constant equal to 5 sec and coefficient of amplification equal to 25.

We consider the equations of direct control

$$\left.\begin{aligned} T_1 \frac{dx_1}{dt} + x_1 &= -k_1 x_1 , \\ T_2'^2 \frac{d^2 x_2}{dt^2} + T_k' \frac{dx_2}{dt} + x_2 &= k_2 x_1 , \end{aligned}\right\} \tag{3.4}$$

where x_1 is the coordinate of the object, and x_2 the coordinate of the controller, T_1 and k_1 the time constant and coefficient of amplification of the object, T_2' and k_2 the time constant and coefficient of amplification of the controller, and T_k is its damping time constant.

We may now formulate the given problems as follows: given the values $T_2' = 0{\cdot}45$; $T_k = 5$; $k_2 = 25$; to find the region of stability in the T_1, k_1 plane

For equation (3.4) the characteristic equation can be written

$$(T_1 p + 1)(T_2'^2 p^2 + T_k p + 1) + k_1 k_2 = 0$$

or, with the given values of T_2', T_k and k_2

$$(T_1 p + 1)(0{\cdot}2\, p^2 + 5p + 1) + 25k_1 = 0 ,$$

or

$$0{\cdot}2 \mu p^3 + 5 \mu p^2 + \mu p + 0{\cdot}2p^2 + 5p + 1 + 25\eta = 0 , \tag{3.5}$$

where $\mu = T_1$ and $\eta = k_1$.

Proceeding to the construction of the D-partition of the equation (3.5), for the parameters μ and η, we substitute $i\omega$ for p:

$$-0{\cdot}2 \mu i \omega^3 - 5\mu\omega^2 + \mu i \omega - 0{\cdot}2\omega^2 + 5i\omega + 1 + 25\eta = 0 .$$

Separating real and imaginary parts and equating them separately to zero, we obtain

$$\left.\begin{aligned} -5\mu\omega^2 + 25\eta + (1 - 0{\cdot}2\omega^2) &= 0, \\ \mu\omega(1 - 0{\cdot}2\omega^2) + \eta{\cdot}0 + 5\omega &= 0. \end{aligned}\right\}$$

In the given case

$$S_1(\omega) = -5\omega^2 ; \quad S_2(\omega) = \omega(1 - 0{\cdot}2\,\omega^2) ; \quad Q_1(\omega) = 25 ;$$
$$Q_2(\omega) = 0 ; \quad R_1(\omega) = 1 - 0{\cdot}2\omega^2 ; \quad R_2(\omega) = 5\omega ;$$

$$\Delta = \begin{vmatrix} S_1(\omega) & Q_1(\omega) \\ S_2(\omega) & Q_2(\omega) \end{vmatrix} = \begin{vmatrix} -5\omega^2 & 25 \\ \omega(1-0{\cdot}2\omega^2) & 0 \end{vmatrix} = -25\omega(1-0{\cdot}2\omega^2);$$

$$\Delta_1 = \begin{vmatrix} -R_1(\omega) & Q_1(\omega) \\ -R_2(\omega) & Q_2(\omega) \end{vmatrix} = \begin{vmatrix} -(1-0{\cdot}2\omega^2) & 25 \\ -5\omega & 0 \end{vmatrix} = 125\omega;$$

$$\Delta_2 = \begin{vmatrix} S_1(\omega) & -R_1(\omega) \\ S_2(\omega) & -R_2(\omega) \end{vmatrix} = \begin{vmatrix} -5\omega^2 & -(1-0{\cdot}2\,\omega^2) \\ \omega(1-0{\cdot}2\omega^2) & -5\omega \end{vmatrix} =$$

$$= 25\omega^2 + \omega(1-0{\cdot}2\omega^2)^2.$$

Hence

$$\mu = \frac{\Delta_1}{\Delta} = -\frac{5}{1-0{\cdot}2\omega^2}, \quad \eta = \frac{\Delta_2}{\Delta} = -\frac{25\omega^2 + (1-0{\cdot}2\omega^2)^2}{25(1-0{\cdot}2\omega^2)}.$$

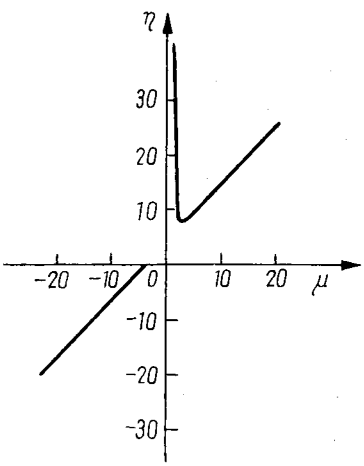


FIG. 115

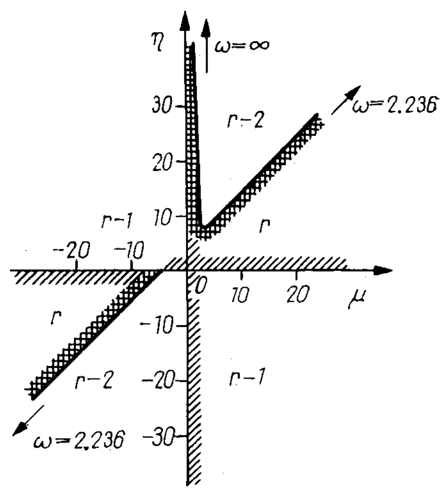


FIG. 116

Giving ω various values from 0 to ∞ we draw the curve shown in Fig. 115 through these points.

We note that Δ becomes zero for $\omega = 0$ and for $\omega = \sqrt{5} \approx 2{\cdot}236$ but for $\omega = \sqrt{5}$ neither Δ_1 nor Δ_2 becomes zero. For $0 < \omega < 2{\cdot}236$ the determinant $\Delta < 0$, and for $\omega > 2{\cdot}236$ $\Delta > 0$.

Moving along the curve from the point $\omega = 0$ to the point $\omega = 2{\cdot}236$, we shade it doubly (Fig. 116) on the right (since $\Delta < 0$) and later, for $\omega > 2{\cdot}236$, on the left (since $\Delta > 0$).

In the given case there are two singular straight lines: one of them corresponds to $\omega = 0$, the other to $\omega = \infty$, since in this case the coefficient of the highest order term of the characteristic equation contains μ.

Equating to zero the free term of the equation, we find $25\eta = -1$ or $\eta = -0{\cdot}04$. Hence, the line $\eta = -0{\cdot}04$* is the first singular line corresponding to $\omega = 0$.

Finally, equating the coefficient of p^3 in equation (3.5) to zero, we find the equation of the second line: $\mu = 0$.

Hence, the η-axis is the second singular line.

In Fig. 116 we show these lines are shaded according to the above rules.

Let the point $\eta = 10$, $\mu = 20$ represent an equation having r negative roots. Figure 116 shows the marking of all the regions for this case.

It is obvious that the plane contains a region of stability, since there are always objects which the given controller can stably control, whatever the values of the parameters T_7^2, T_k and k_2, provided that they are all different from zero.

The region (r) has the greatest mark in Fig. 116, and it is therefore the required region of stability.

We may test the correctness of this reasoning by putting $\eta = -0{\cdot}04$; $\mu = 1$. Then the equation (3.5) becomes

$$0{\cdot}2p^3 + 5{\cdot}2p^2 + 6p = 0$$

and has roots

$$p_1 = 0\,;\; p_{2,3} = \frac{-\,5{\cdot}2 \pm \sqrt{5{\cdot}2^2 - 4.6.0{\cdot}2}}{2.0{\cdot}2}\,.$$

Of these three roots, two are situated to the left of the imaginary axis, and one on it. The point $\eta = -0{\cdot}04$; $\mu = 1$ lies on the singular line and crossing to the region (r) leaves the singular line on the shaded side.

Hence the regions (r) correspond to equations with all three roots lying to the left of the imaginary axis.

In order to find out the value of r, we can proceed as follows: in (3.5) let us put $\mu = 0$; $\eta = 0{\cdot}2$. Then (3.5) reduces to the quadratic equation

$$0{\cdot}2p^2 = 5p + 6 = 0\,.$$

Both of its roots are negative.

The point $\mu = 0$, $\eta = 0{\cdot}2$ lies on the singular line $\mu = 0$, corresponding to $\omega = \infty$. When we leave this line on the shaded side the number of roots with a negative real part is increased by one and therefore $r = 3$.

(e) *Criteria for stability. Necessary condition for stability. For stability it is necessary that all the coefficients in the characteristic equation have the same sign*

From Bézout's theorem, any equation can be factorized into factors of the form $(p - p_k)$ where p_k are the roots:

$$a_0\,p^n + a_1\,p^{n-1} + \ldots + a_n = a_0\,(p - p_1)\,(p - p_2)\ldots\,(p - p_n)\,.$$

* In Fig. 116 the line $\eta = 0{\cdot}04$ is combined with the line $\eta = 0$. The left branch of the curve in the first quadrant is now drawn to scale.

But if the real parts of all the roots lie to the left of the imaginary axis, then each bracket contains a positive term.* Hence $a_1, a_2, \ldots, a_n$ have the same sign.

The A. V. Mikhailov criterion. Let us turn to the characteristic equation

$$D(p) = a_0 p^n + a_1 p^{n-1} + \ldots + a_{n-1} p + a_n = 0. \tag{3.6}$$

Let us introduce the parameter λ and consider the more general class of equations

$$D(p) = \lambda ,$$

of which equation (3.6) most frequently occurs (where $\lambda = 0$).

Let us put $p = i\omega$ in equation (3.6) and construct the D-partition for the parameter λ (Fig. 117).

The only possible region of stability is the region A, shaded in Fig. 117*a*. If the point $\lambda = 0$ lies inside this region and the latter is the region of stability, then all roots of the equation lie to the left of the imaginary axis and the system is stable.

We can prove a rule which enables us to establish whether the region A is the region of stability, without having recourse to shading the curve.

Let us consider the case $\lambda = 0$, i.e. the equation (3.6). Let the degree of this equation be equal to n and let $p_1, p_2, \ldots, p_n$ — be its roots. Then

$$D(p) = a_0 (p - p_1)(p - p_2) \ldots (p - p_n) .$$

Putting $i\omega$ in place of p:

$$D(i\omega) = a_0 (i\omega - p_1)(i\omega - p_2) \ldots (i\omega - p_n) .$$

* For the real root $p_k = -|p_k|$ this is obvious:

$$(p - p_k) = (p + | p_k|).$$

In the case of a pair of complex conjugate roots we have:

$$(p - p_k)(p - p_{k+1}) = [p - (a_k + i\beta_k)][p - (a_k - i\beta_k)] = \\ = p^2 - 2a_k p + (a_k^2 + \beta_k^2) ,$$

and if $a_k < 0$, then $a_k = -|a_k|$ and

$$(p - p_k)(p - p_{k+1}) = p^2 + 2 | a_k | p + (a_k^2 + \beta_k^2).$$

If, in the p-plane, we mark the point p_k, then the difference $i\omega - p_k$ corresponds to the vector joining the point p_k to the point $i\omega$ on the imaginary axis (Fig. 117*b*). Let us go from $\omega = -\infty$ to $\omega = +\infty$. In the case when p_k lies to the left of the imaginary axis, the argument of this vector $(i\omega - p_k)$, will be increased by π (Fig. 117*c*). Therefore, if all the roots $p_1, p_2, \ldots, p_n$ lie to the left of the imaginary axis, then when ω changes from $-\infty$ to $+\infty$ the argument of the vector $D\,(i\,\omega)$ (sometimes called the *characteristic vector)*, increases by $\pi\,n$, which represents the sum of the increments in the arguments of all n vectors $i\,\omega - p_k$. And if r roots lie to the right of

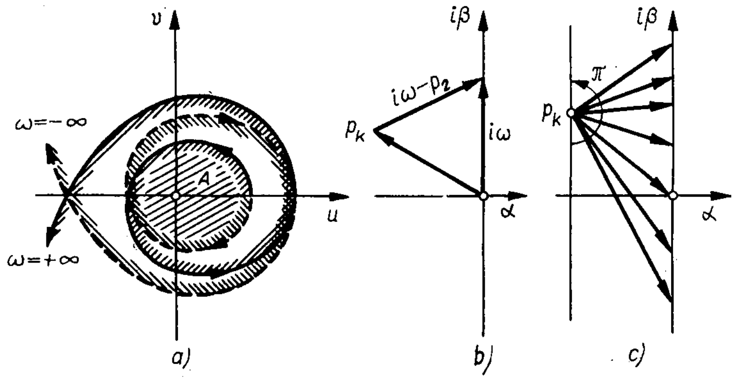


FIG. 117

the imaginary axis, and $n - r$ roots to the left of it, then the tota increment in the argument is equal to $\{(n - r) - r\}\,\pi$.

Let us restrict the change in ω from 0 to $+\infty$. Then the change in the argument of the characteristic vector $D(i\omega)$ will be half as great as it is when ω changes from $-\infty$ to $+\infty$, and will be equal tol $n\dfrac{\pi}{2}$ when the system is stable. We note that as ω increases the argument of all the vectors $(i\,\omega - p_k)$ for which p_k lies to the left of the imaginary axis and, of course, the argument of the vector $D(i\,\omega)$ of the stable system, grows monotonically.

Thus, the region A will be the region of stability if, when for $\omega = 0$ we start on the real axis to the right of the coordinate origin, and for ω increasing to $+\infty$ the boundary of the D-partition for the given λ passes through n quadrants (where n is the degree of

the given characteristic equation) in such a way that the radius vector always rotates in an anti-clockwise direction.

Let us consider now the left-hand side of any characteristic equation $D(p)$. Let us replace p by $i\,\omega$ and separate real and imaginary parts:

$$D(i\,\omega) = u(\omega) + iv(\omega)\,.$$

Let us put $u(\omega)$ along the x-axis, and $v(\omega)$ along the y-axis. Then, calculating the quantities u and v for values of ω, we find points in this plane. Letting ω go from zero to infinity, we can trace a curve in this plane, the *hodograph of the characteristic equation* of the system. The radius vector drawn from the origin of coordinates to this point, as we have already shown, is sometimes called the characteristic vector of the system. The argument of this vector will be, as usual, equal to the angle between the positive direction of the u-axis and the radius-vector, reckoned anticlockwise.

We may then formulate the following criterion for stability (the A. V. Mikhailov criterion).

In order that a system be stable, it is necessary and sufficient that the modulus of the characteristic vector be different from zero for any $\omega(0 \leqslant \omega \leqslant \infty)$ *and that the argument of this vector be equal to zero for* $\omega = 0$, *and, further, that as* ω *increases monotonically from* 0 *to* $+\infty$, *it should increase monotonically from* 0 *to* $\frac{\pi}{2}n$, *where* n *is the degree of characteristic equation.*

In other words, the hodograph of the characteristic equation of a stable system may only be drawn in the following way: beginning for $\omega = 0$ on the semi-axis u^+, it must then encompass the origin of coordinates and subsequently intersect the semi-axes v^+, u^-, v^-, u^+, v^+, u^- and so on, until it has traversed n quadrants. In the nth quadrant the hodograph goes off to infinity.

The hodograph is like this only when

$$u(0) > 0\,;\ v(0) = 0;$$

$$\left[\frac{dv}{d\omega}\right]_{\omega=0} > 0$$

and when the zeroes of the function $v(\omega)$ alternate with the zeroes of the function $u(\omega)$ (Fig. 118).

If the hodograph of the characteristic equation is different in any way, for example if it crosses the same semi-axis twice in succession, then the system will be unstable.

The Routh—Hurwitz criterion. We consider below several different formulations of this criterion. Not differing from each other in principle, they differ in form and each of them has its most appropriate uses.

Let us put $p = i\omega$ in the characteristic equation $D(p) = 0$ of the system, and separate real and imaginary parts:

$$D(i\omega) = u(\omega) + iv(\omega) .$$

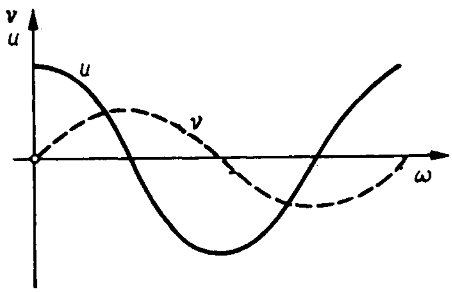


FIG. 118

We recall that in a stable system the real roots of the equations $u(\omega) = 0$ and $v(\omega) = 0$ alternate (Fig. 118).

On the D-partition boundary $u(\omega) = 0$ and $v(\omega) = 0$ simultaneously. The condition for this is that a determinant of the nth order, called the highest Hurwitz determinant, shall be equal to zero. It is formed according to the scheme:

$$\Delta_n = \begin{vmatrix} a_1 & a_3 & a_5 & a_7 & \dots & 0 \\ a_0 & a_2 & a_4 & a_6 & \dots & 0 \\ 0 & a_1 & a_3 & a_5 & \dots & 0 \\ 0 & a_0 & a_2 & a_4 & \dots & 0 \\ 0 & 0 & a_1 & a_3 & \dots & 0 \\ 0 & 0 & a_0 & a_2 & \dots & 0 \\ 0 & 0 & 0 & a_1 & \dots & 0 \\ 0 & 0 & 0 & a_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & a_n \end{vmatrix} = 0,$$

where the a_k are the coefficients in the characteristic equation:

$$D(p) = a_0 p^n + a_1 p^{n-1} + a_2 p^{n-2} + \ldots + a_n.$$

The equation $\Delta_n = 0$ determines a surface in the coefficient space which includes the boundary of the D-partition.

The region of stability is found from the condition that the determinant Δ_n itself and all its diagonal minors

$$\Delta_1 = a_1\,; \quad \Delta_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad \Delta_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}; \quad \ldots$$

shall be positive.

It has been shown that when all the coefficients of the characteristic equation are positive ($a_0 > 0, a_1 > 0, \ldots, a_n > 0$) from the fact that all the determinants Δ_1, Δ_3, Δ_5 with odd indices are positive, it follows that the determinants Δ_2, Δ_4, Δ_6, ... with even indices are also positive, and vice versa. Hence when the necessary conditions for stability are fulfilled, i.e. $a_0 > 0, a_1 > 0, \ldots, a_n > 0$, the necessary and sufficient conditions for stability may be formulated as follows:

In order that all roots of the characteristic equation

$$a_0 p^n + a_1 p^{n-1} + \ldots + a_n = 0$$

with positive coefficients ($a_0 > 0$, $a_1 > 0, \ldots, a_n > 0$) *shall have negative real parts, it is necessary and sufficient that among the determinants*

$$\Delta_1, \Delta_2, \ldots, \Delta_n$$

all determinants with even (or all determinants with odd) indices shall be positive.

Let us consider for example the system whose characteristic equation is

$$a_0 p^3 + a_1 p^2 + a_2 p + a_3 = 0.$$

The stability condition, according to Hurwitz's criterion, will be that the inequalities.

$$\Delta_3 = \begin{vmatrix} a_1 & a_3 & 0 \\ a_0 & a_2 & 0 \\ 0 & a_1 & 0_3 \end{vmatrix} > 0\,; \quad \Delta_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix} > 0\,; \quad \Delta_1 = a_1 > 0,$$

shall be satisfied, or, after expansion of the determinants $\Delta_2 = a_1 a_2 - a_0 a_3 > 0$ and $\Delta_3 = a_3 \Delta_2 > 0$.

If all the coefficients a_0, a_1, a_2 and a_3 are positive, then we require only that

$$\Delta_2 = a_1 a_2 - a_0 a_3 > 0 .$$

The evaluation of the determinants $\Delta_1, \Delta_2, \ldots, \Delta_n$ by means of decomposing them according to the elements of any column or row is usually tedious. This is why it is convenient not to compute these determinants, but to bring the determinant Δ_n to diagonal form.

From the theory of determinants we know that if we subtract from all the elements of any row of the determinant a multiple of the elements of any other row, then neither the value of the determinant, nor that of its diagonal minors, will be changed.

Let us consider the last but one Hurwitz determinant* Δ_{n-1}:

$$\Delta_{n-1} = \begin{vmatrix} a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & a_{13} & a_{15} & \cdots \\ a_0 & a_2 & a_4 & a_6 & a_8 & a_{10} & a_{12} & a_{14} & \cdots \\ 0 & a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & a_{13} & \cdots \\ 0 & a_0 & a_2 & a_4 & a_6 & a_8 & a_{10} & a_{12} & \cdots \\ 0 & 0 & a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & \cdots \\ 0 & 0 & a_0 & a_2 & a_4 & a_6 & a_8 & a_{10} & \cdots \\ 0 & 0 & 0 & a_1 & a_3 & a_5 & a_7 & a_9 & \cdots \\ 0 & 0 & 0 & a_0 & a_2 & a_4 & a_6 & a_8 & \cdots \end{vmatrix}$$

All the other Hurwitz determinants are the diagonal minors of Δ_{n-1}.

Let us form the ratio of the first elements of the first two rows:

$$\lambda_1 = \frac{a_0}{a_1} .$$

We subtract from all the elements of the second row the corresponding elements of the first row, first multiplied by $\lambda_1 = \dfrac{a_0}{a_1}$.

* Due to the fact that $\Delta_n = a_n \Delta_{n-1}$ and $a_n > 0$, the inequality $\Delta_{n-1} > 0$ implies $\Delta_n > 0$.

As a result the second row is replaced by the row

$$0\ \bar{a}_2\ \bar{a}_4\ \bar{a}_6\ \bar{a}_8\ \ldots ,$$

where

$$\bar{a}_2 = a_2 - \frac{a_0}{a_1} a_3, \quad \bar{a}_4 = a_4 - \frac{a_0}{a_1} a_5$$

and so on.

In an exactly similar way, the fourth, sixth, eighth, etc. rows may be transformed, no new calculations being required since in Δ_{n-1} each subsequent pair of rows repeats the first and second rows, displaced by one column. We may thus immediately replace a_0 by 0, a_2 by $\bar{a}_2$, a_4 by $\bar{a}_4$ and so on, everywhere.

As a result we arrive at the determinant

$$\begin{vmatrix} a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & a_{13} & a_{15} & \ldots \\ 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \bar{a}_{12} & \bar{a}_{14} & \ldots \\ 0 & a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & a_{13} & \ldots \\ 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \bar{a}_{12} & \ldots \\ 0 & 0 & a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & \ldots \\ 0 & 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \ldots \\ 0 & 0 & 0 & a_1 & a_3 & a_5 & a_7 & a_9 & \ldots \\ 0 & 0 & 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{vmatrix}$$

We now subtract from each element of the third row of this determinant the element of the second row lying above it, first multiplied by $\lambda_2 = \frac{a_1}{a_2}$, i. e. we replace the third row $0\ a_1\ a_3\ a_5\ a_7\ a_{11}\ a_{13} \ldots$ by the row

$$0\ 0\ \bar{\bar{a}}_3\ \bar{\bar{a}}_5\ \bar{\bar{a}}_7\ \bar{\bar{a}}_9\ \bar{\bar{a}}_{11}\ \bar{\bar{a}}_{13} ,$$

where

$$\bar{\bar{a}}_3 = a_3 - \frac{a_1}{\bar{a}_2} \bar{a}_4, \quad \bar{\bar{a}}_5 = a_5 - \frac{a_1}{\bar{a}_2} \bar{a}_6$$

and so on.

Just as before we repeat this substitution in the fifth, seventh, ninth, rows, etc., i.e. in the third row and in all lower rows we replace a_1 by 0, a_3 by $\bar{\bar{a}}_3$, a_5 by $\bar{\bar{a}}_5$ and so on.

As a result of this transformation we obtain the determinant:

$$\begin{vmatrix} a_1 & a_3 & a_5 & a_7 & a_9 & a_{11} & a_{13} & a_{15} & \dots \\ 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \bar{a}_{12} & \bar{a}_{14} & \dots \\ 0 & 0 & \bar{\bar{a}}_3 & \bar{\bar{a}}_5 & \bar{\bar{a}}_7 & \bar{\bar{a}}_9 & \bar{\bar{a}}_{11} & \bar{\bar{a}}_{13} & \dots \\ 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \bar{a}_{12} & \dots \\ 0 & 0 & 0 & \bar{\bar{a}}_5 & \bar{\bar{a}}_7 & \bar{\bar{a}}_9 & \bar{a}_{11} & \bar{\bar{a}}_{13} & \dots \\ 0 & 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \bar{a}_{10} & \dots \\ 0 & 0 & 0 & \bar{\bar{a}}_3 & \bar{\bar{a}}_5 & \bar{\bar{a}}_7 & \bar{\bar{a}}_9 & a_{11} & \dots \\ 0 & 0 & 0 & 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \bar{a}_8 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{vmatrix}.$$

Continuing the process, we are left with a determinant of the same order:

$$\begin{vmatrix} a_1 & a_3 & a_3 & a_7 & \dots \\ 0 & \bar{a}_2 & \bar{a}_4 & \bar{a}_6 & \dots \\ 0 & 0 & \bar{\bar{a}}_3 & \bar{\bar{a}}_5 & \dots \\ 0 & 0 & 0 & \bar{\bar{\bar{a}}}_4 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix}$$

in which all the elements lying to the left of the principal diagonal are zero and the elements of the main diagonal are

$$a_1, \bar{a}_2, \bar{\bar{a}}_3, \bar{\bar{\bar{a}}}_4, \dots$$

The value of this determinant is equal to the product of its diagonal elements

$$\delta_{n-1} = a_1 . \bar{a}_2 . \bar{\bar{a}}_3 . \bar{\bar{\bar{a}}}_4 \cdot \dots,$$

and any diagonal minor of the kth order is equal to the product of the k diagonal elements of this minor.

The determinant δ_{n-1} and its diagonal minors coincide respectively with the Hurwitz determinant Δ_{n-1} and its diagonal minors, since the transformation from the determinant Δ_{n-1} to δ_{n-1} changes neither the value of the determinant nor the value of any of its diagonal minors.

Thus

$$\Delta_1 = a_1,$$
$$\Delta_2 = a_1 . \bar{a}_2,$$
$$\Delta_3 = a_1 . \bar{a}_2 . \bar{\bar{a}}_3$$

and so on, and the Hurwitz condition $\Delta_1 > 0$, $\Delta_2 > 0$, $\Delta_3 > 0, \dots$ reduces to the condition $a_1 > 0$, $\bar{a}_2 > 0$, $\bar{\bar{a}}_3 > 0$.

Therefore the Routh—Hurwitz criterion may be formulated as follows:

We form the determinant Δ_{n-1} and using transformations which alter neither value of the determinant nor the value of its diagonal minors, we reduce it to diagonal form. In order that the system shall be stable it is necessary and sufficient that all the elements situated on the main diagonal of this determinant shall be positive.*

We note that when using the Routh—Hurwitz criterion it is important only to know the sign of the determinant and of its diagonal minors, and not their value. And the sign of the determinant is not altered if we multiply all elements of any row or column by an arbitrary positive number M. We can use this in order to reduce all the elements of a row or column by a common multiple.

We may arrive at the determinant δ_{n-1} in another way. From the coefficients of the characteristic equation

$$a_0 p^n + a_1 p^{n-1} + \dots + a_{n-1} p + a_n = 0 \tag{3.6}$$

we construct the Routh table:

No. of row	No of column				
	1	2	3	4	5
1	a_0	a_2	a_4	a_6	...
2	a_1	a_3	a_5	a_7	...
3	a_{31}	a_{32}	a_{33}	a_{34}	...
4	a_{41}	a_{42}	a_{43}	...	...
5	a_{51}	a_{52}	...	...	...
6	...	...	...	...	...

where

$$a_{31} = \frac{a_1 a_2 - a_0 a_3}{a_1}, \quad a_{41} = \frac{a_{31} a_3 - a_1 a_{32}}{a_{31}}, \quad a_{51} = \frac{a_{41} a_{32} - a_{31} a_{42}}{a_{41}},$$

$$a_{32} = \frac{a_1 a_4 - a_0 a_5}{a_1}, \quad a_{42} = \frac{a_{31} a_5 - a_1 a_{33}}{a_{31}}, \quad a_{52} = \frac{a_{41} a_{33} - a_{31} a_{43}}{a_{41}}.$$

$$a_{33} = \frac{a_1 a_6 - a_0 a_7}{a_1}, \quad a_{43} = \frac{a_{31} a_7 - a_1 a_{34}}{a_{31}}$$

$$a_{34} = \frac{a_1 a_8 - a_0 a_9}{a_1}.$$

* It is assumed that the necessary condition of stability, i.e. that all the coefficients including a_n are positive, is satisfied.

The elements of the first row are all the coefficients with even indices, and the elements of the second row those with odd indices. The elements of the third row are obtained by the cross-product of the elements of the first two rows, subsequently divided by the first element of the previous row. The elements of each subsequent row are formed in the same way from the two preceding rows. Then the Routh—Hurwitz criterion may be formulated as follows:

In order that the real part of all roots of the equation (3.6) *shall be negative, it is necessary and sufficient that all the elements in the first column of the above table, i.e.*

$$a_0$$

$$a_1$$

$$a_{31} = \frac{a_1 a_2 - a_0 a_3}{a_1}$$

$$a_{41} = \frac{a_{31} a_3 - a_1 a_{32}}{a_{31}}$$

shall be different from zero and of the same sign.

If a_0 is negative, then it can always be made positive, by multiplying both sides of (3.6) by -1. Therefore with $a_0 > 0$ the system is stable if all the elements of the first column of the table are positive.

The formation of the Routh table can be stopped as soon as the first element of any row turns out to be negative or equal to zero.

Numerical Example. Let us find out whether the system with characteristic equation

$$p^4 + 8p^3 + 18p^2 + 16p + 5 = 0\,.$$

is stable or not. We construct the Routh table:

No. of row	No. of column 1	2	3
1	1	18	5
2	8	16	0
3	$\frac{8\cdot 18 - 1\cdot 16}{8} = 16$	$\frac{8.5 - 1.0}{8} = 5$	$\frac{8.0 - 1.0}{8} = 0$
4	$\frac{16.16 - 8.5}{16} = 13\cdot 5$	$\frac{16.0 - 8.0}{16} = 0$	0
5	$\frac{13\cdot 5 - 16.0}{13\cdot 5} = 5$	$\frac{13\cdot 5.5 - 16.0}{13\cdot 5} = 0$	0

The elements of the first column are equal to 1, 8, 16, 13.5 and 5; they are all positive, and hence the system is stable.

We notice at once that the operations which it is necessary to perform on the coefficients a in order to form the Routh table are exactly the same as those which it is necessary to perform on these coefficients when transforming the Hurwitz determinant to diagonal form.

It is sometimes convenient to perform these operations without constructing tables. It will be shown later that the given equation

$$a_0 p^n + a_1 p^{n-1} + \ldots + a_n = 0$$

may be transformed to the equation

$$\begin{aligned}\left(a_0 - \frac{a_0}{a_1} a_1\right) p^n + a_1 p^{n-1} + \left(a_2 - \frac{a_0}{a_1} a_3\right) p^{n-2} + \\ + a_3 p^{n-3} + \left(a_4 - \frac{a_0}{a_1} a_5\right) p^{n-4} + \ldots + \ldots = 0.\end{aligned} \tag{3.7}$$

and that the $(n-1)$ roots of this new equation are in the same situation, with respect to the imaginary axis, as are $(n-1)$ of the roots of the original equation; the remaining root of the original equation being removed to either $\pm\infty$.

We write this new equation as

$$\bar{a}_0 p^m + \bar{a}_1 p^{m-1} + \bar{a}_2 p^{m-2} + \ldots = 0, \tag{3.8}$$

where

$$m = n - 1$$

and

$$\begin{aligned}\bar{a}_0 &= a_1, \\ \bar{a}_1 &= a_2 - \frac{a_0}{a_1} a_3, \\ \bar{a}_2 &= a_3, \\ \bar{a}_3 &= a_4 - \frac{a_0}{a_1} a_5\end{aligned}$$

and so on. We again apply the same transformation to (3.8)

$$\left(\bar{a}_0 - \frac{\bar{a}_0}{\bar{a}_1} \bar{a}_1\right) p^m + \bar{a}_1 p^{m-1} + \left(\bar{a}_2 - \frac{\bar{a}_0}{\bar{a}_1} \bar{a}_3\right) p^{m-2} + \ldots = 0,$$

i.e. we proceed to the equation

$$\bar{\bar{a}}_0 p^r + \bar{\bar{a}}_1 p^{r-1} + \ldots = 0\,, \tag{3.9}$$

where

$$r = m - 1 = n - 2$$

and

$$\bar{\bar{a}}_0 = \bar{a}_1, \bar{\bar{a}}_1 = \bar{a}_2 - \frac{\bar{a}_0}{\bar{a}_1}\,\bar{a}_3$$

and so on.

Continuing in this way to successively lower the degree of the characteristic equation, we finally arrive at an equation of zero degree. The stability criterion may now be stated.

If all the ratios

$$\frac{a_0}{a_1},\quad \frac{\bar{a}_0}{\bar{a}_1},\quad \frac{\bar{\bar{a}}_0}{\bar{\bar{a}}_1},\ldots$$

are positive, then the roots of the characteristic equation lie to the left of the imaginary axis and the system is stable.

But if all the coefficients of the characteristic equation are positive, then $\dfrac{a_0}{a_1} > 0$; therefore $\dfrac{\bar{a}_0}{\bar{a}_1} > 0$, if $\bar{a}_1 > 0$, since $\bar{a}_0 = a_1$; further $\dfrac{\bar{\bar{a}}_0}{\bar{\bar{a}}_1} >$ > 0 if $\bar{\bar{a}}_1 > 0$ since $\bar{\bar{a}}_0 = a_1$, and so on.

Thus the conditions for stability reduce to the form

$$a_1 > 0,\ \bar{a}_1 = a_2 - \frac{a_0}{a_1}\,a_3 > 0\,,$$

$$\bar{\bar{a}}_1 = a_2 - \frac{a_1}{\bar{a}_1}\left(a_4 - \frac{a_0}{a_1}\,a_5\right) > 0\,.$$

Example. Let us apply this sequence of operations to the equation

$$p^6 + 6p^5 + 21p^4 + 44p^3 + 62p^2 + 52p + 24 = 0\,.$$

This leads to the equation

$$6p^5 + \left(21 - \frac{1}{6}\,44\right)p^4 + 44p^3 + \left(62 - \frac{1}{6}\,52\right)p^2 + 52p + 24 = 0$$

or

$$6p^5 + 13{\cdot}67p^4 + 44p^3 + 53{\cdot}34p^2 + 52p + 24 = 0\,.$$

By the same method we obtain

$$13.67p^4 + \left(44 - \frac{6}{13{\cdot}67}\,53.34\right)p^3 + 53.34p^2 + \left(52 - \frac{5}{13{\cdot}67} \times 24\right)p + 24 = 0$$

or

$$13{\cdot}67p^4 + 20{\cdot}6p^3 + 53{\cdot}34p^2 + 41{\cdot}45p + 24 = 0\,.$$

Once again decreasing the degree of the equation we have

$$20{\cdot}6\, p^3 + \left(53{\cdot}34 - \frac{13{\cdot}67}{20{\cdot}6} \times 41{\cdot}45\right) p^2 + 41{\cdot}45\, p + 24 = 0$$

or

$$20{\cdot}6p^3 + 25{\cdot}9p^2 + 41{\cdot}45p \;\; +24 = 0.$$

Using the same transformation

$$25{\cdot}9\, p^2 + \left(41{\cdot}45 - \frac{20{\cdot}6 \times 24}{25{\cdot}9}\right) p + 24 = 0$$

or

$$25{\cdot}9p^2 + 22{\cdot}25p + 24 = 0.$$

The system is stable, since

$$\frac{1}{6} > 0, \quad \frac{6}{13{\cdot}67} > 0, \quad \frac{13{\cdot}67}{20{\cdot}6} > 0, \quad \frac{20{\cdot}6}{25{\cdot}9} > 0,$$

and the quadratic equation we have obtained has roots with a negative real part, since all its coefficients are positive.

We see at once that this sequence of operations on the coefficients is the same as that performed in constructing the Routh table or in reducing the Hurwitz determinant to diagonal form.

Let us proceed to a proof of the Routh—Hurwitz criterion.

We described above both the reduction which successively lowers the degree of the characteristic equation, and the criterion of Routh—Hurwitz, which is equivalent to it in the sense that when using the reduction and when using the Routh—Hurwitz criterion in its usual form the same operations must be carried out on the coefficients of the characteristic equation. Therefore in order to prove the Routh—Hurwitz criterion it is sufficient to prove this reduction.

The essence of the proof consists in the following.

The use of the reduction lowers the degree of the characteristic equation by one. It is proved (see below) that as a result of this reduction the arrangement of the roots relative to the imaginary axis is not changed, but one of the roots is taken to ∞ without crossing the imaginary axis in the root plane.

By repeating the reduction n times in sequence all the roots are taken to infinity. If they are taken to $-\infty$, i.e. if they move in the left-hand half-plane, then all roots of the initial characteristic equation also lay in the left-hand half-plane. It is proved that the condition that all $\frac{a_0}{a_1} > 0$ ensures that all the roots are sent to $-\infty$.

We prove the reduction for even n. For odd n it is proved similarly.

Let $S(p)$ be a polynomial in p with real coefficients and let $S(i\,\omega) = u(\omega) + iv(\omega)$. We show first of all that as a result of using the reduction the polynomial $S(p)$ is transformed to the polynomial $S_1(p)$ such that

$$S_1(i\,\omega) = [u(\omega) + r\,\omega\,v(\omega)] + iv(\omega)\,, \tag{3.10}$$

where

$$r = \frac{a_0}{a_1}\,.$$

Let

$$S\,(p) = a_0\,p^n + a_1\,p^{n-1} + \ldots + a_n. \tag{3.11}$$

Then for even n

$$u\,(\omega) = \gamma a_0\,\omega^n - \gamma a_2\,\omega^{n-2} + \ldots, \tag{3.12}$$

$$v\,(\omega) = -\,\gamma a_1\,\omega^{n-1} + \gamma a_3\,\omega^{n-3} - \ldots, \tag{3.13}$$

where $\gamma = +1$ if $n = 4, 8, 12, \ldots$, and $\gamma = -1$ if $n = 2, 6, 10, \ldots$

Hence

$$u\,(\omega) + r\omega v\,(\omega) = \gamma\omega^n\,(a_0 - ra_1) - \gamma\omega^{n-2}\,(a_2 - ra_3) + \ldots \tag{3.14}$$

But

$$S_1(p) = (a_0 - ra_1)\,p^n + a_1\,p^{n-1} + (a_2 - ra_3)\,p^{n-2} + \ldots$$

Therefore $\mathrm{Re}\,S_1(i\,\omega)$ is determined by the equation (3.14), and $\mathrm{Im}\;S_1(i\omega)$ by (3.13).

Hence (3.10) is proved.

We now construct the D-partition of $S_1(p)$ for the parameter r. To do this we solve $S_1(i\,\omega) = 0$ with respect to r,

$$r = -\,\frac{u\,(\omega)}{\omega v(\omega)} - i\,\frac{1}{\omega}\,.$$

The point $r = 0$, corresponds to the initial polynomial $S(p)$ and the point $r = \frac{a_0}{a_1}$ to the polynomial $S_1(p)$ obtained after reduction.

We consider now the shape of the D-partition boundary for r. We note first of all that it intersects the real axis only for $\omega = \infty$, and for any $0 \leq \omega < +\infty$ it lies beneath it. For $\omega = \infty$ we obtain

$$r = -\,\frac{\gamma a_0}{-\gamma a_1} = \frac{a_0}{a_1}\,.$$

For $\omega = 0$ we have $r = -\infty$.

The sign of the imaginary part of r is always the reverse of the sign of ω.

Thus the D-partition boundary for $\frac{a_0}{a_1} > 0$ is as in Fig. 119a, and for $\frac{a_0}{a_1} < 0$ as in Fig. 119b.

In the first case (Fig. 119*a*), going from the point $r = \frac{a_0}{a_1}$ corresponding to the equation $S_1(p) = 0$ to the point $r = 0$, corresponding to the initial equation $S(p) = 0$ we move one root over to the left of the imaginary axis, since we leave the D-curve on the shaded side. The polynomial $S(p)$ for $r = 0$ has one root more to the left of the imaginary axis than for $r = \frac{a_0}{a_1}$.

In the case of Fig. 119*b* on the contrary, when going from $r = \frac{a_0}{a_1}$ to $r = 0$ the number or roots lying to the right of the imaginary axis is increased by one.

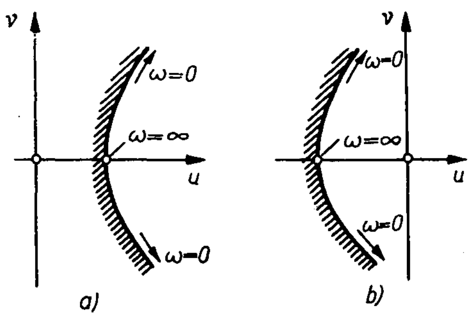


FIG. 119

Thus, for even n, the root tends to ∞ on the left of the imaginary axis, if $\frac{a_0}{a_1} > 0$.

Hence, the above reduction transforms any equation of the nth degree to another equation $S_1(p) = 0$ of degree $n - 1$. $n - 1$ roots of the original equation are, in relation to the imaginary axis, just the same as all $n - 1$ roots of the equation $S_1(p)$, and the last, nth, root lies to the left of the imaginary axis, since in the reduction with $\frac{a_0}{a_1} > 0$, it is taken to $-\infty$.

2. An Estimate of the Stability of a System from the Frequency Characteristics

(a) *The first amplitude-phase criterion of stability*

Let the system have the characteristic euqation

$$D(p) + K(p) = 0 \,. \tag{3.15}$$

Let us consider the more general equation

$$D(p) - \lambda\, K(p) = 0 \,, \tag{3.15}$$

where λ is complex.

Obviously (3.15) can be obtained from (3.16) by putting $\lambda = -1$.

We find the region of stability for (3.16) for the parameter λ by constructing the D-partition in the λ-plane.

We solve the equation (3.16) with respect to λ, and obtain

$$\lambda = \frac{D(p)}{K(p)}.$$

Then we put $p = i\omega$, so that

$$\lambda = \frac{D(i\omega)}{K(i\omega)} = u(\omega) + iv(\omega).$$

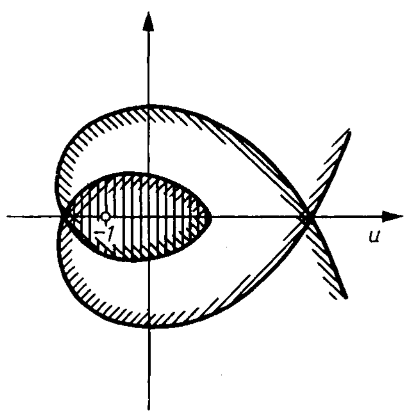


FIG. 120

Putting in $u(\omega)$ and $v(\omega)$ values of ω from $-\infty$ to $+\infty$ we construct the boundary of the D-partition (Fig. 120).

The amplitude-phase characteristic of the first kind, with the addition of its mirror reflection in the real axis, is the boundary of the D-partition. If such a characteristic is given as the initial material for the calculation (See Chapter II) there is no need, therefore, to construct the boundary of the D-partition.

By shading, we determine the region which corresponds to the polynomials having the greatest number of roots to the left of the imaginary axis (this is done in Fig. 120).

To determine the conditions under which this region corresponds to the region of stability, let us take the boundary point $\lambda = 0$ which corresponds to the equation

$$D(p) = 0 \; . \tag{3.17}$$

Then, in the λ-plane (Fig. 120) equation (3.17) corresponds to the point $\lambda = 0$, and this is the characteristic for an open system, while (3.15), which is the characteristic for a closed system, corresponds to the point $\lambda = -1$.

If the open system is stable the point $\lambda = 0$ belongs to the region of stability in the λ-plane. The given closed system is stable if the point $\lambda = -1$ belongs to this region; thus, for a system which is stable

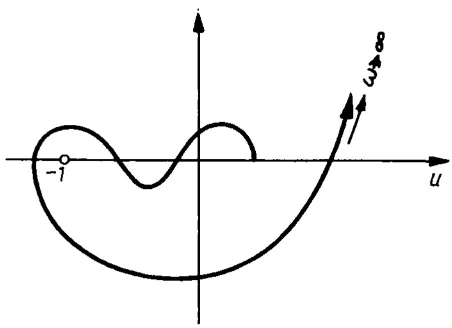


FIG. 121

in the open state, we may formulate the following criterion for stability:

In order that a system which is stable in the open state be stable also in the closed state, it is necessary and sufficient that the amplitude-phase characteristic of the first kind does not intersect the segment of the real axis between 0 and -1 or else intersects this segment from above downwards ("descent") and from below upwards ("ascent") the same number of times (Fig. 121).

This criterion may be generalized to cover the case when the equation

$$D(p) = 0$$

has $2r + \varepsilon$ roots (where $r = 0, 1, 2, \ldots$, and $\varepsilon = 1$ if the number of roots is odd and 0 if it is even) situated to the right of the imaginary axis or on it, assuming that there are no repeated roots on the axis.

We will move along the real axis in the λ-plane from the point $\lambda = 0$ to the point $\lambda = -1$. The number of roots to the left of the imaginary axis increases by two when we "ascend", corresponding to $\omega > 0$, and decreases by two when we "descend"*.

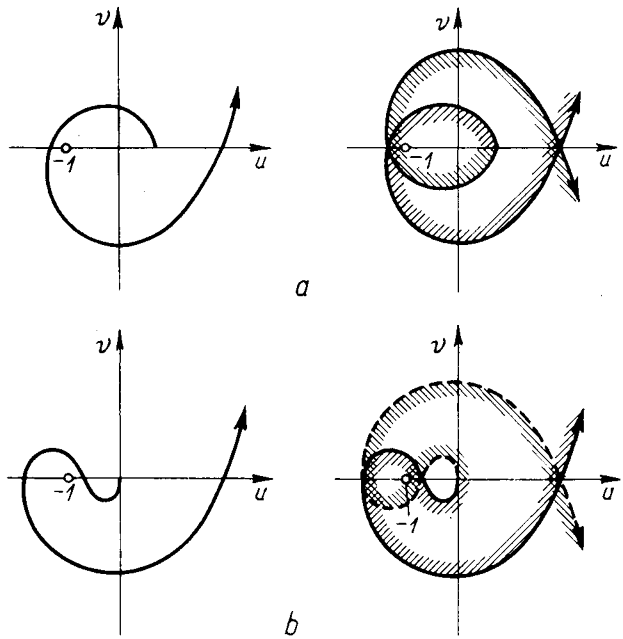


FIG. 122

Similarly, an "ascent" for $\omega = 0$ in the segment $-1 < u \leqslant 0$ enables us to transfer one root to the left.

In order that the system with the characteristic equation

$$D(p) + K(p) = 0\,,$$

shall be stable when the total number of roots of the equation $D(p) = 0$ lying to the right of the imaginary axis and on it (but with no repeated points on it) is equal to $2r + \varepsilon$ (where r is any whole positive number

* The number of roots to the left of the imaginary axis changes by two, since after adding to the amplitude-phase characteristic its mirror reflection in the real axis, all the points of intersection with the u-axis for $\omega < 0$ are points of self-intersection of the curve.

and $\varepsilon = 1$ for an odd and $\varepsilon = 0$ for an even number of roots), it is necessary and sufficient that :

(1) in the segment $-1 < u \leqslant 0$ the number of "ascents" of the amplitude-phase characteristic of the first kind for $\omega > 0$ exceeds the number of "descents" by r, and

(2) when $\varepsilon = 1$ the point $\omega = 0$ corresponds to an ascent in this same segment of the axis, and when $\varepsilon = 0$ or in the presence of a zero root, to a descent.

Figure 122 shows the shape of some amplitude-phase characteristics of the first kind when the equation $D(p) = 0$ has no roots to the right of the imaginary axis or on it (Fig. 122*a*), or has one root to the right of the imaginary axis and one zero root (Fig. 122*b*), and the closed system is stable. For the sake of clarity the right-hand sides of Fig. 122*a* and *b* show the shaded characteristic together with its mirror reflection in the imaginary axis.

(b) *Second amplitude-phase criterion for stability*

If we are given an amplitude-phase characteristic of the second kind, instead of the first kind (i.e. if we are given the hodograph of $\dfrac{D(i\omega)}{K(i\omega)}$ instead of $\dfrac{K(i\omega)}{D(i\omega)}$ then instead of equation (3.16) we must consider the equation

$$K(p) - \lambda\, D(p) = 0\,.$$

Then the amplitude-phase characteristic of the second kind $\dfrac{K(i\,\omega)}{D(i\,\omega)}$ is the D-partitio n boundary.

The point $\lambda = -1$ corresponds to the equation $D(p) + K(p) = 0$, i.e. the characteristic equation of a closed system. The point $\lambda = -\infty$ corresponds to the equation $D(p) = 0$, i.e. the characteristic equation of an open system. Because of this we must now move from the point $\lambda = -\infty$ to $\lambda = -1$ instead of from $\lambda = 0$ to $\lambda = -1$.

All the reasoning given in the construction of the first amplitude phase criterion is still valid, except that now the number of ascents must not exceed the number of descents, but, on the contrary, the number of descents must exceed the number of ascents by r. Thus the amplitude-phase criterion for stability may be formulated as follows:

In order that the system with the characteristic equation $D(p) + K(p) = 0$ shall be stable in the case when the total number of roots of the equation $D(p) = 0$ lying to the right of the imaginary axis and on it (but with no repeated roots on it) is equal to $2r + \varepsilon$ (where $\varepsilon = 1$ or 0) it is necessary and sufficient that

(1) on the segment $-\infty \leqslant u < -1$ the number of "descents" for $\omega > 0$ of the amplitude-phase charactersitic of the second kind shall exceed the number of "ascents" by r, and that

(2) when $\varepsilon = 1$ the point $\omega = 0$ corresponds to a "descent" in this segment of the u-axis, and when $\varepsilon \mp 0$ or when there is a zero root, to an "ascent".

By way of example, Fig. 123 shows the shape of the amplitude-phase characteristic of the second kind for the same cases as in Fig. 122.

(c) *Estimating the stability from the logarithmic characteristic*

We showed in Chapter II that in some case the construction of the logarithmic characteristic is exceptionally simple. In these cases we use the logarithmic characteristic of the open system as initial data for the calculation. To judge the stability from these characteristics, we can use them to construct the ordinary amplitude-phase characteristic and then use the amplitude-phase criterion for stability. We need not, however, carry out such a construction but can adapt the criterion to the peculiarities of logarithmic characteristics.

We restrict ourselves to the case when the open system is stable. In accordance with the second amplitude-phase criterion for stability,

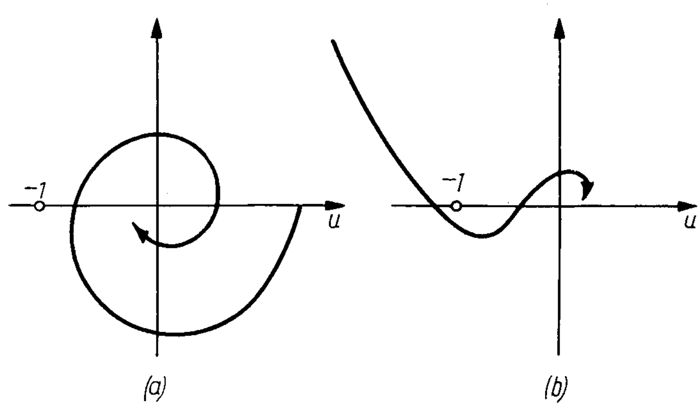


FIG. 123

for the stability of the closed system it is necessary and sufficient in this case that the amplitude-phase characteristic of the second kind descends and ascends in the section $-\infty \leqslant u < -1$ of the real axis the same number of times. This condition is fulfilled if, for all ω for which the argument φ of the characteristic vector is equal to $\pm k\,\pi (k = 1, 3, \ldots)$ and the modulus of the vector is greater than unity, the signs of $\dfrac{d\varphi}{d\omega}$ alternate. But $20 \log 1 = 0$. Hence, in order to find out whether the closed system is stable in the case when the open system is, it is necessary to mark all values of ω for which the phase characteristic intersects the line $\pm k\,\pi (k = 1, 3, \ldots)$ and to find the sign of $\dfrac{d\varphi}{d\omega}$ for those ω for which the logarithmic amplitude characteristic lies above the frequency axis.

(d) *The use of amplitude-phase characteristics in estimating the stability of systems containing elements with time delays or with distributed parameters*

The concept of a time delay. Everywhere in the preceding account it was supposed that the action of the output coordinate of one stage on the input of the next stage was realized instantaneously. In practice, there are often occasions when such an assumption cannot be made, since the time required for transmission of the signal is not small enough to be ignored.

The calculation of the delay time constant, if it is at all considerable,* or the calculation of the complete picture of the wave process in pipes can be important in the analysis of the stability of the system.

The characteristic equation of a system with a time delay. Let us remove from the system the element which introduces the time delay. Then we obtain an ordinary open linear system. Its equation is

$$\begin{aligned} a_0 \frac{d^n x_{\text{out}}}{dt^n} + a_1 \frac{d^{n-1} x_{\text{out}}}{dt^{n-1}} + \ldots + a_n x_{\text{out}} = \\ b_0 \frac{d^m x_{\text{in}}}{dt^m} + b_1 \frac{d^{m-1} x_{\text{in}}}{dt^{m-1}} + \ldots + b_m x_{\text{in}}. \end{aligned} \tag{3.18}$$

* We will later define more exactly what value of τ is considerable (see p. 202).

The system in reality is closed across the element introducing the delay, and therefore

$$x_{\text{in}}(t) = -x_{\text{out}}(t-\tau). \tag{3.19}$$

Putting (3.19) in equation (3.18) we obtain for $x_{\text{out}}(t)$ a differential equation with a "delay argument":

$$\begin{aligned} a_0 \frac{\mathrm{d}^n x_{\text{out}}(t)}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1} x_{\text{out}}(t)}{\mathrm{d}t^{n-1}} + \ldots + a_n x_{\text{out}}(t) = \\ = -\left[b_0 \frac{\mathrm{d}^m x_{\text{out}}(t-\tau)}{\mathrm{d}t^m} + b_1 \frac{\mathrm{d}^{m-1} x_{\text{out}}(t-\tau)}{\mathrm{d}t^{m-1}} + \ldots \right. \\ \left. \ldots + b_m x_{\text{out}}(t-\tau)\right]. \end{aligned} \tag{3.20}$$

As in the usual linear equation, we look for an integral of equation (3.20) in the form

$$x_{\text{out}}(t) = Ce^{pt}. \tag{3.21}$$

We note that in this case

$$a_0 \frac{\mathrm{d}^n x_{\text{out}}}{\mathrm{d}t^n} + \ldots + a_n x_{\text{out}} = D(p)\, x_{\text{out}} = D(p)\, Ce^{pt}, \tag{3.22}$$

where

$$D(p) = a_0 p^n + a_1 p^{n-1} + \ldots + a_n,$$

and the function $x_{\text{out}}(t-\tau)$ is transformed to

$$x_{\text{out}}(t-\tau) = Ce^{p(t-\tau)}.$$

Then

$$\frac{\mathrm{d}x_{\text{out}}(t-\tau)}{dt} = Cpe^{p(t-\tau)}$$

and generally

$$\frac{\mathrm{d}^m x_{\text{out}}(t-\tau)}{\mathrm{d}t^m} = Cp^m e^{p(t-\tau)}$$

Therefore, using (3.21)

$$\begin{aligned} b_0 \frac{d^m x_{\text{out}}(t+\tau)}{dt^m} b_1 + \frac{d^{m-1} x_{\text{out}}(t-\tau)}{dt^{m-1}} + \ldots + b_m x_{\text{out}}(t+\tau) = \\ = K(p)\, Ce^{p(t-\tau)} = K(p)\, Ce^{pt} e^{-p\tau} \end{aligned} \tag{3.23}$$

where

$$K(p) = b_0 p^m + b_1 p^{m-1} + \ldots + b_m$$

Putting the last expression and (3.22) in (3.20) and dividing the right and left-hand sides by Ce^{pt}, we find:

$$D(p)\,K(p)\,e^{-p\tau} = 0 \tag{3.24}$$

The roots of this transcendental equation are those values of p for which the function (3.21) is an integral of the equation (3.20)

In contrast to an algebraic equation, the transcendental equation (3.24) can have an infinite number of roots, and the corresponding complete integral of equation (3.20) is equal to

$$x_{\text{out}} = \sum_{j=1}^{j=\infty} C_j\, e^{p_j t}, \tag{3.25}$$

where p_j are the roots of equation (3.24) and C_j are constants determined by the boundary conditions.

The system is obviously stable if all the terms in (3.25) tend to zero as $t \to \infty$, i.e. if all the roots of (3.24) lie to the left of the imaginary axis in the plane of the roots.*

Equation (3.24) plays exactly the same role with respect to the initial equation (3.20) as the algebraic characteristic equation plays for the usual linear differential equation.

For this reason equation (3.24) is called the *characteristic equation* for the initial equation (3.20).

We note that for $\tau = 0$ (3.24) becomes

$$D(p) + K(p) = 0\,,$$

which is the ordinary characteristic equation of the system without delay.

For the transcendental equation (3.24) it is now necessary to consider the same problem of the distribution of the roots relative to the imaginary axis, which we considered for a polynomial above.

The amplitude-phase criterion of stability for a transcendental characteristic equation. Let us return to equation (3.24). We shall consider the more general equation

$$-\lambda\, D(p) + K(p)e^{-\tau p} = 0\,, \tag{3.26}$$

* The reverse statement, that the system is unstable if among the roots of equation (3.24) there is any one root with a positive real part, is considerably less trivial. It will be proved for the present only with some supplementary conditions on the boundary conditions of the problem.

where λ is any complex number. This equation becomes equation (3.24) for $\lambda = -1$.

We construct the D-partition of (3.26) for the parameter λ:

$$\lambda(p) = \frac{K(p)}{D(p)} e^{-\tau p}.$$

After substituting $i\omega$ for p we obtain:

$$\lambda(i\omega) = \frac{K(i\omega)}{D(i\omega)} e^{-\tau i\omega}.$$

For $\lambda = -\infty$ (3.26) becomes $D(p) = 0$, and the distribution of the roots of this equation is independent of τ. We first put $\tau = 0$. Let us find how many times, in moving along the real axis in the λ-plane from the point $\lambda = -\infty$ to the point $\lambda = -1$, it is necessary to cross the D-partition boundary on the shaded side and on the unshaded side, in order that the point $\lambda = -1$ shall belong to the region of stability. It will then be necessary to cross the D-partition boundary exactly the same number of times in the same way for any $\tau > 0$ in order that the point $\lambda = -1$ shall belong to the region of stability.

For the case $\tau = 0$ the required number of intersections and their correct direction are ensured if the conditions of the amplitude-phase criterion of stability are fulfilled. This criterion is true, therefore, not only for polynomials, but also for the transcendental characteristic equation (3.24).

But if, for $\tau = 0$, the equation of the amplitude-phase characteristic is

$$\lambda(i\omega) = \frac{K(i\omega)}{D(i\omega)}, \tag{3.27}$$

then for $\tau > 0$ this equation is

$$\lambda(i\omega) = \frac{K(i\omega)}{D(i\omega)} e^{-i\tau\omega}. \tag{3.28}$$

The stability of systems with delays. In order to construct the characteristics (3.28) it is first necessary to construct the hodograph (3.27) and then to displace all the points so that the radius-vector of each point is unchanged but its argument is decreased by $\tau\omega$.

Suppose, first, that for $\tau = 0$ the system is stable and has the amplitude-phase characteristic of the second kind given in Fig. 124.

Further, let τ be small, but not zero. Then each radius vector turns through an angle $\tau\omega$ in a clockwise direction. Thus for example

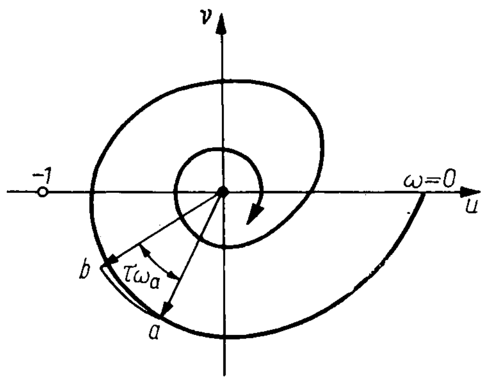


FIG. 124

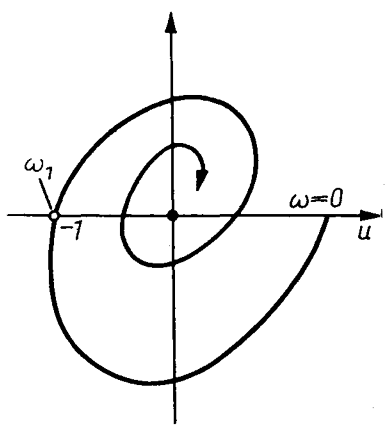


FIG. 125

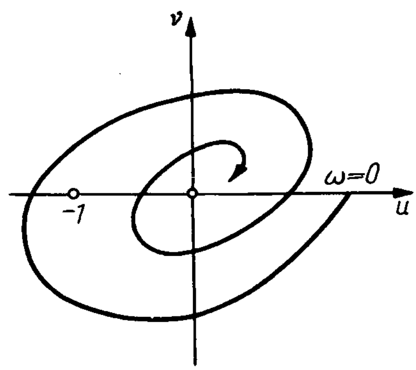


FIG. 126

the point a (Fig. 124) is displaced to the point b. It may happen that after such a deformation the characteristic does not contain the point $\lambda = -1$.

We then increase τ and once more deform the hodograph. It may happen that for some $\tau = \tau_1$ the characteristic passes through the point $\lambda = -1$ (Fig. 125), and for $\tau = \tau_1 + \varepsilon$ surrounds it (Fig. 126).

To find this value of τ_1 we draw an arc of unit radius as far as its first intersection with the characteristic. Let the point of intersection correspond to $\omega = \omega_1$, and let τ_1 be the ratio of this arc (in angular measure) to ω_1. Then for $\tau = \tau_1$ the characteristic passes through the point $\lambda = -1$, and the system is on the boundary of stability — it generates undamped oscillations with a frequency ω_1 (Fig. 125).

When τ is increased the system becomes unstable (Fig. 126). Stability is once again established (Fig. 127) if τ crosses to the value τ_2,

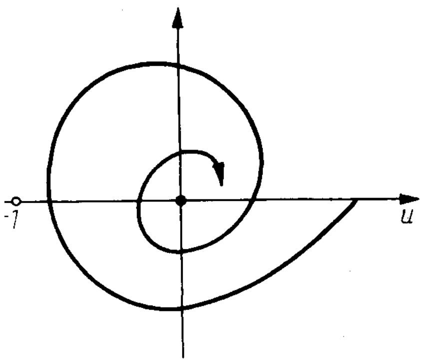


Fig. 127

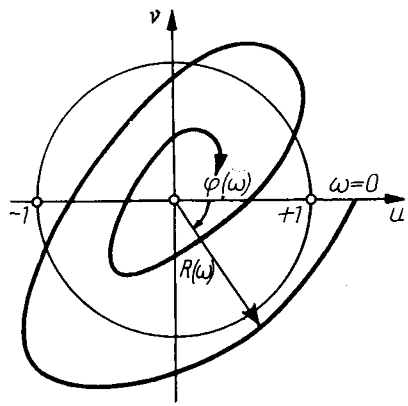


Fig. 128

corresponding to the second intersection by the characteristic of the point $\lambda = -1$.

Thus, as τ increases from zero to infinity, the system can either be always stable, or else regions of stability and instability can alternate.

To determine these regions we construct a circle of unit radius, and determine all the points of intersection of the characteristic with this circle (Fig. 128). These points are determined by the equations

$$R(\omega) = 1\,, \qquad \varphi(\omega) - \tau\omega = -\pi(2k + 1)\,, \tag{3.29}$$

where $k = 0, 1, 2, \ldots$

Solving the first of the equations (3.29) for ω, we find the values of the frequencies ω_j at the points of intersection with the unit circle. From the second equation we obtain

$$\tau_j = \frac{\varphi(\omega_j)}{\omega_j} + \frac{\pi(2k + 1)}{\omega_j}\,. \tag{3.30}$$

These values of τ_j divide the region of possible values of τ into stable and unstable sections, i.e. they isolate the region of stability for the parameter τ (Fig. 129). The system is stable for all τ, only if the whole characteristic lies inside the unit circle, i.e. if $\left|\frac{K(i\omega)}{D(i\omega)}\right| < 1$ for $0 \leqslant \omega < \infty$. But this case is impractical since it requires that $K(0)$ should be excessively small.

Remarks on the calculation of the wave processes in pipes. It is sometimes impossible to ignore the presence of reflected waves in pipes and the phenomenon of hydraulic thrust or other wave phenomena have to be taken into account.

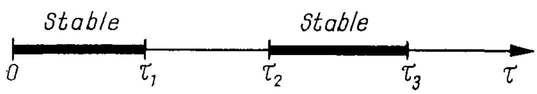


FIG. 129

In this case the total differential equation must be supplemented by partial differential equations, and their corresponding boundary conditions.

If these equations and the boundary conditions are linear, the characteristic equation of such a system reduces to

$$A(p)\, e^{p\tau_1} + B(p)\, e^{-p\tau_1} = 0\,. \tag{3.31}$$

It is not difficult to extend the amplitude-phase criterion for stability to equation (3.31) as well.

Let us consider the more general equation

$$-\lambda A(p)\, e^{p\tau_1} + B(p)\, e^{-p\tau_1} = 0\,, \tag{3.32}$$

which becomes (3.31) when $\lambda = -1$.

From equation (3.32) we have

$$\lambda(p) = \frac{B(p)}{A(p)}\, e^{-2p\tau_1}$$

or

$$\lambda(i\omega) = \frac{B(i\omega)}{A(i\omega)}\, e^{-2i\omega\tau_1}. \tag{3.33}$$

If the hodograph of the vector (3.33) is called the amplitude-phase characteristic of the system having characteristic equation (3.31) then the amplitude-phase criterion for stability and all the above statements concerning delay remain valid for this case also. In this case. for the system corresponding to $\tau = 0$ whose characteristic is constructed in order to determine its intersection with the unit circle we have to consider not $\frac{K(i\omega)}{D(i\omega)}$, but $\frac{B(i\omega)}{A(i\omega)}$, and instead of the distribution of the roots of the open system $D(p) = 0$ we must take as the equivalent the distribution of the roots of the equation $A(p) = 0$. In addition, in this case, $\tau = 2\tau_1$.

3. General Properties of Some Classes of Systems of Automatic Control, Connected with the Conditions of their Stability

The criteria of stability may be used not only in order to determine the conditions for stability in each concrete case, but also to study the general properties connected with stability for whole classes of control systems. Knowing these properties we may immediately make a number of inferences about the stability of a system from its scheme, without using the criteria of stability.

(a) *The conditions for the existence of the region of stability in the parameter space (the conditions for structural stability)*

The concept of the conditions for structural stability. Let us suppose that any system is given by its structural scheme. It is known what stages it contains and what kind of connexions exist between these stages. The values of the time constants or the coefficients of amplification of the stages are, however, not known. The aggregate of the positive numerical values of all the time constants and other coefficients which it is necessary to know in order to calculate the value of the coefficients of the characteristic equation when we know the structural scheme, are called the *system parameters.*

A change in the non-zero parameters, i.e. in the positive values of the time constants, the coefficients of amplification and so on, does not change the structural scheme of the system in any way.

The structural scheme of the system is usually predetermined by the properties of the controlled object and by the chosen scheme of the controller. In a number of cases the system proves to be unstable for all values of the parameters and we may only make it stable by changing the structure of the system. Systems of this kind are called structurally-stable systems, which can be made stable by the choice of parameters.

Examples of structurally-unstable systems are given in Fig. 103a and b, and an example of a structurally-stable system is given in Fig.

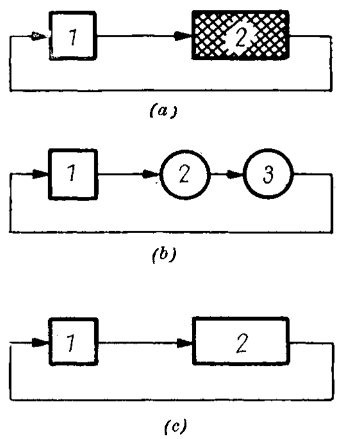


Fig. 130

130c.* For the system shown in Fig. 130a the characteristic equation can be put in the form

$$(T_1p + 1)(T_2^2p^2 + 1) + k_1k_2 = 0$$

or

$$T_1T_2^2p^3 + T_2^2p^2 + T_1p + (k_1k_2 + 1) = 0\,.$$

The Routh—Hurwitz criteria require that the inequality

$$T_1T_2^2 - T_1T_2^2\,(k_1k_2 + 1) > 0\,,$$

shall be satisfied. This reduces to

$$-T_1T_2^2\,k_1k_2 > 0$$

* From Fig. 130 onwards the conditional notation for stages, introduced in Chapter II (p. 102), is used.

and hence cannot be satisfied if T_1, T_2, k_1 and k_2, are positive numbers. The system is unstable for any positive values of T_1, T_2, k_1 and k_2, i. e. it is structurally unstable.

For the system shown in Fig. 130b the characteristic equation can be put in the form

$$(T_1 p + 1)\, pp + k_1 \bar{k}_2 \bar{k}_3 = 0$$

or

$$T_1 p^3 + p^2 + k_1 \bar{k}_2 \bar{k}_3 = 0 .$$

For any values of T_1, $k_1 \bar{k}_2$ and $\bar{k}_3$ the system is unstable, since its characteristic equation does not contain a first degree term in p. Hence, the system is structurally-unstable.

Let us consider, finally, the system shown in Fig. 130c. Its characteristic equation is

$$(T_1 p + 1)\,(T_2^2 p^2 + T_{2k} p + 1) + k_1 k_2 = 0$$

or

$$T_1 T_2^2 p^3 + (T_1 T_{2k} T_2^2) p^2 + (T_1 + T_{2k}) p + (k_1 k_2 + 1) = 0 .$$

The Routh—Hurwitz criteria reduce to the inequality

$$(T_1 T_{2k} + T_2^2)\,(T_1 + T_{2k}) - T_1 T_2^2\,(k_1 k_2 + 1) > 0 ,$$

which can always be satisfied, for example, by taking the values of k_1 and k_2 sufficiently small. The system, therefore, is structurally-stable.

The problem consists in deciding from the form of the structural scheme whether the system is structurally-stable or structurally-unstable, i. e whether the parameter space of the system contains a region of stability or not.

The conditions for the structural stability of single-loop systems without derivative action. We first consider single-loop systems with the characteristic equation

$$D(p) + K = 0 ,$$

where

$$D(p) = \Pi d_j\,(p) ,$$

and $d_j\,(p)$ are polynomials of the zero, first or second degree of the form

$$Tp + 1, \quad Tp - 1, \quad p,$$

$$T^2 p^2 + T_k\, p + 1, \quad T^2 p^2 + 1 .$$

We denote by q the number of astatic stages (having $d(p) = p$), by t the number of unstable stages (i.e. with $d(p) = Tp - 1$), by r the number of conservative stages (with $d(p) = T^2p^2 + 1$)* in the system, and by n the degree of the polynomial $D(p)$.

The conditions for the structural stability of such a system are determined by the following theorem:

In order that a single-loop system shall be structurally-stable it is necessary and sufficient that, simultaneously, the inequalities

(a) $q + t < 2$ *and* (b) $n > 4r$ *shall be satisfied*

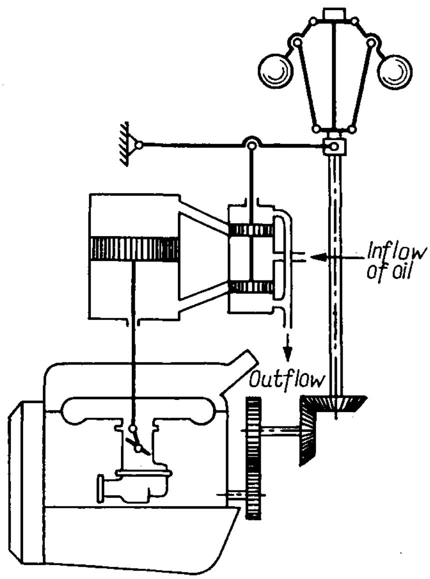


FIG. 131

As an example let us consider the assembly of the indirect control of an internal combustion engine using a controller without feedback (Fig. 131).

* The calculation of the conservative stages is important for two reasons. Firstly, it is often necessary to ensure stability for limited small damping in the oscillatory stage. In this case we neglect damping, i. e. replace the oscillatory stage by a conservative stage, and require that the system shall be stable in this case too. Secondly, two astatic stages following one another and closed by feedback transmitted from the output of the second stage to the input of the first, form a conservative stage (see footnote on p. 210). Such a combination of astatic stages is often encountered in control systems, since it allows us to use only one feedback when there are two cascades of astatic amplifiers.

The engine represents a single-loop stage if it possesses inherent regulation, an astatic stage if it does not and an unstable stage if the inherent regulation is negative. The static sensor represents either an oscillatory stage (if viscous friction is taken into account), or a conservative stage (if not). The servomotor without feedback represents an astatic stage.

Table IV gives the various possible structural schemes, depending on the properties of the engine and of the controlled object.

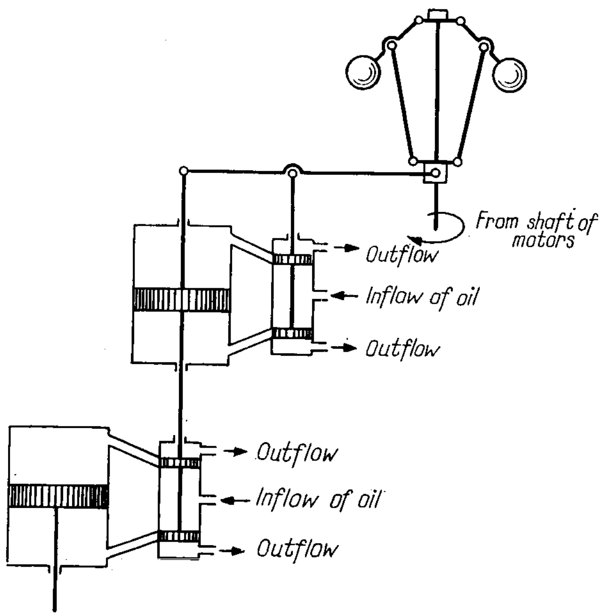


FIG. 132

We see that out of six cases in only one can the system be made stable: when the engine possesses positive inherent regulation and friction is taken into account in the sensor element (scheme *a*).

Let us add one more servomotor possessing feedback, this being equivalent to adding one single-capacitance stage to the assembly (Fig. 132).

The structural schemes for the same six systems are shown in Table V.

The system can now be made stable for the engine which possesses positive inherent regulation both with friction in the sensor, and without it (Table Va and b), but as before it is not possible to produce stability in any of the other cases.

In place of the additional servomotor included in the scheme of Fig. 131, let us shunt the servomotor in the scheme with feedback.

TABLE IV

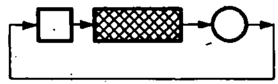
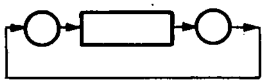
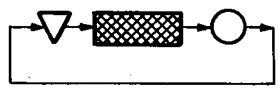
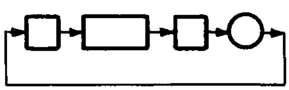
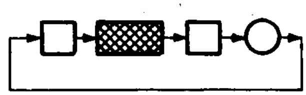
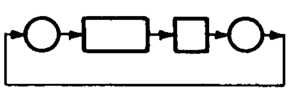
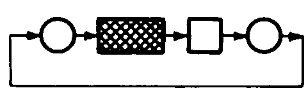
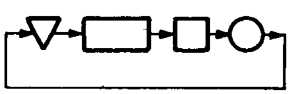
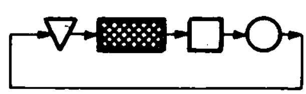
Inherent regulation	Sensor: With friction	Sensor: Without friction
Motor with positive inherent regulation	 a)	 b)
Motor without inherent regulation	 c)	 d)
Motor with negative inherent regulation	 e)	 f)

TABLE V

Inherent regulation	Sensor: With friction	Sensor: Without friction
Motor with positive inherent regulation	 a)	 b)
Motor without inherent regulation	 c)	 d)
Motor with negative inherent regulation	 e)	 f)

Then the system will be structurally-stable for any engine, provided there is friction in the sensor (Table VIa, c and e), but is still structurally-unstable for all engines without friction in the sensor (Table VIb, d and f).

Let us turn to the scheme of two-stage amplification, but let us add feedback in the second servomotor (Fig. 133).

Now (Table VII) the system is structurally-stable in all six cases, i.e. both for any engine and with or without friction in the sensor.

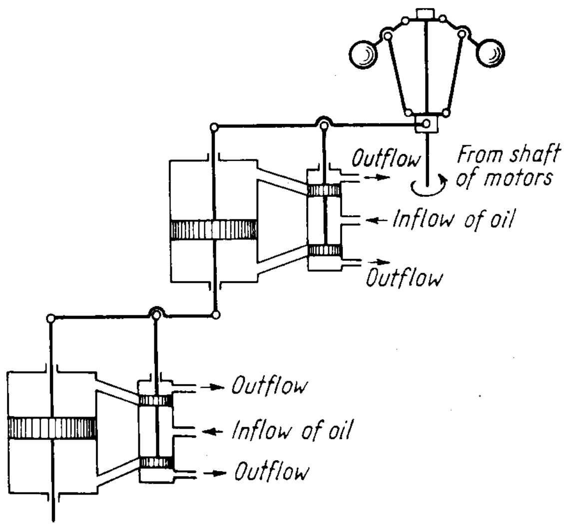


FIG. 133

The conditions for structural stability in single-loop systems with derivative action. If positive first-derivative action is present in a single-loop system, the characteristic equation of the system takes the form

$$D(p) + Rp + K = 0\,,$$

where R is a positive number.

If, in addition, there is positive second derivative action, or if first derivative actions exist at two points in the circuit, then the characteristic equation of the system reduces to the form

$$D(p) + Mp^2 + Rp + K = 0,$$

where M and R are positive numbers.

Bearing in mind the general case of the system containing any

TABLE VI

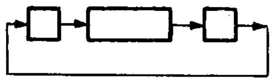
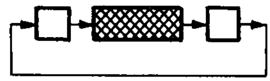
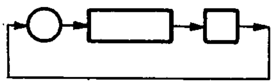
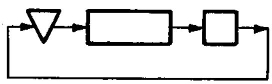
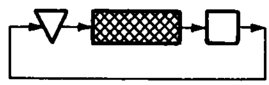
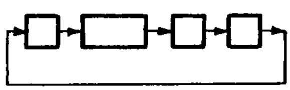
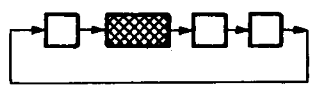
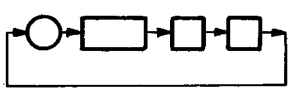
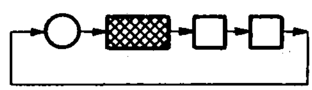
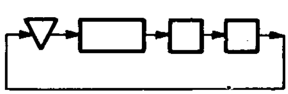
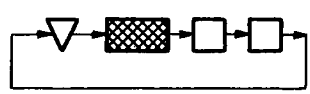
Inherent regulation	Sensor	
	With friction	Without friction
Motor with positive inherent regulation	 a)	 b)
Motor without inherent regulation	 c)	 d)
Motor with negative inherent regulation	 e)	 f)

TABLE VII

Inherent regulation	Sensor	
	With friction	Without friction
Motor with positive inherent regulation	 a)	 b)
Motor without inherent regulation	 c)	 d)
Motor with negative inherent regulation	 e)	 f)

number of first derivative actions or of both first and second derivative actions, let us consider the characteristic equation

$$D(p) + K(p) = 0. \tag{3.34}$$

Here $K(p)$ is the product of factors of the form $(R_i p + K_i)$ and $(M_i p^2 + R_i p + K_i)$ and R_i, M_i and K_i are positive, non-zero, numbers.* There can be any number of such factors provided that m, the degree of $K(p)$, does not exceed n, the degree of $D(p)$.

With these assumptions, all the roots of the polynomial $K(p)$ lie to the left of the imaginary axis, or, as is sometimes said, $K(p)$ is a Hurwitzian polynomial.

Just as before, we suppose that $D(p)$ contains only factors of the form p (the number of such factors being q), $Tp - 1$ (their number being t), $T'^2 p^2 + 1$ (their number being r), $Tp + 1$ and $T'^2 p^2 + T_k p + 1$. In addition to the earlier notation we put $\mu = q + t + 2r$, $\varrho =$ the integral part of the fraction $\mu/2$, m is the degree of the polynomial $K(p)$, and $N = n + m$.

The conditions for the structural stability of systems containing the derivative actions given above are determined by the following theorem :

In order that the system with characteristic equation (3.34), *in which $D(p)$ is the product of factors of the form p, $Tp + 1$, $T'^2 p^2 + T_k p + 1$, $T'^2 p^2 + 1$ and $Tp - 1$, and $K(p)$ is a Hurwitz polynomial, shall be structurally stable, it is necessary and sufficient that the inequality*

$$m \geqslant q + t - 1 \tag{3.35}$$

be satisfied, and that one of the inequalities in Table VIII, depending upon the values of μ and m, be satisfied.

TABLE VIII

	$m = 0$	$m > 0$ and even	m odd
μ even	$N > 4\varrho$	$N > 4\varrho - 1$	$N > 4\varrho - 2$
μ odd	$N > 4\varrho$	$N > 4\varrho$	$N > 4\varrho + 1$

* This means that there are no negative derivative actions in the circuit, and that positive second derivative actions are introduced at any point of the loop only if there is first derivative action at that point (in the contrary case we would have $M_i > 0$, $R_i = 0$, and this is excluded by the given conditions).

In the simplest particular case, when $K(p) = Rp + K$, i.e. when the system contains only first derivative actions at one point of the circuit, we have $m = 1$ and the inequality (3.35) reduces to $q + t \leqslant 2$.

In this case, m is odd, and we must use the last column of Table VIII.

Three cases are possible: $q + t = 0$, $q + t = 1$ and $q + t = 2$. In the first case $\mu = 2r$, i.e. μ is even, $\varrho = r$ and the second inequality reduces to the form

$$N > 4r - 2$$

or, since $N = n + 1$, to the form

$$n > 4r - 3.$$

In the second case, when $q + t = 1$, we have $\mu = 2r + 1$, i.e. μ is odd, but as before $\varrho = r$. In this case the second inequality reduces to

$$n > 4r.$$

Finally, in the third case, when $q + t = 2$, $\mu = 2r + 2$ and $\varrho = r + 1$ then the second inequality becomes

$$n + 1 > 4(r + 1) - 2$$

or

$$n > 4r + 1.$$

Thus, *if first derivative action is present at one point of the circuit, the necessary and sufficient conditions for structural stability are obtained if the inequality* (3.35) *is replaced by*

$$q + t \leqslant 2,$$

and Table VIII is replaced by the simpler Table IX.

TABLE IX

$q + t$	0	1	2
Inequality	$n > 4r - 3$	$n > 4r$	$n > 4r + 1$

Thus, due to positive first derivative action, the value of $q + t$ may be increased to 2 without disturbing the conditions of structural stability, and when astatic and unstable stages are absent (i.e. for

$q + t = 0$) the number of conservative stages may sometimes be increased without disturbing the stability.

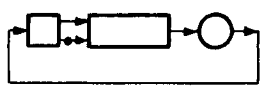
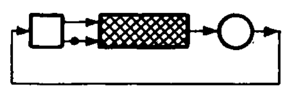
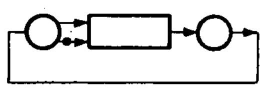
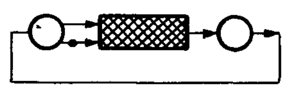
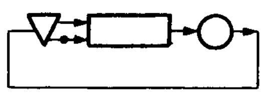
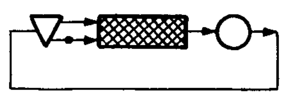
Table X repeats Table IV, but on the added assumption that the system contains positive first derivative action. The bold lines indicate those systems which are known to be stable on the strength of the above theorem of structural stability.

In Table X an arrow with a point $\dashrightarrow$ denotes derivative action.

A comparison between Tables X and IV shows how structural stability is basically favoured by the introduction of derivative action.

In the above theorem it was assumed that $K(p)$ was a Hurwitz polynomial. In order to consider also those cases when the system contains negative derivative action (i.e. when $R_i < 0$ and $M_i < 0$) or when second derivative action is introduced without first derivative action (i.e. when $R_i = 0$ for $M_i \neq 0$) we now lift all restrictions on the polynomial $K(p)$.

TABLE X

Inherent regulation	Sensor	
	With friction	Without friction
Motor with positive inherent regulation	 a)	 b)
Motor without inherent regulation	 c)	 d)
Motor with negative inherent regulation	 e)	 f)

In addition to the condition $m \leqslant n$, we shall assume that the polynomial $D(p)$ consists of factors of the form $Tp + 1$, $T'^2 p^2 + 1$ and $T'^2 p^2 + Tf p + 1$, and contains not more than one factor p (or $Tp - 1$), i.e. we shall assume that $q + t < 2$. The conditions for structural stability in this case are determined by the theorem:

TABLE XI

	$\lambda = 0$	$\lambda > 0$ and even		λ odd	
		(A)	(B)	(A)	(B)
$q + t = 0$	$N > 4r$	$N > 4r - 1$	$N > 4r$	$N > 4r - 2$	$N > 4r - 2$
$q + t = 1$	$N > 4r$	$N > 4r$	$N > 4r$	$N > 4r + 1$	$N > 4r + 2$

In order that the system with characteristic equation

$$D(p) + K(p) = 0$$

for $q + t > 2$ shall be structurally-stable, it is necessary and sufficient that one of the inequalities in Table XI shall be satisfied, this inequality being chosen according to the values of $q + t$ and λ and depending on whether we are considering the normal case (A) or the particular case (B).

In addition to our earlier notation, here λ is the number of roots of the polynomial $K(p)$ lying to the left of the imaginary axis, (A) is the case when one of the factors in $K(p)$ is Hurwitzian, and (B) is the case when $K(p)$ contains no Hurwitzian factors.

As an example we consider a system containing a negative first derivative action at one point in the circuit. In this case $K(p) = -Rp + K$, i.e. $\lambda = 0$ and $N = n + 1$.

Consequently, the condition for structural stability is the inequality

$$n > 4r - 1$$

Comparing this with the inequalities of Table IX, we note that for $q + t = 1$ negative first derivative action can also guarantee structural stability in those cases when positive first derivative action is of no use. Thus, for example, the system with characteristic equation

$$p(Tp + 1)(T'^2 p^2 + 1) + Rp + K = 0$$

is structurally unstable, and the system with characteristic equation

$$p(Tp + 1)(T'^2 p^2 + 1) - Rp + \mathrm{K} = 0$$

is structurally stable.

The structural stability of some non-single-loop systems. The theorems given above may also be used in the analysis of the conditions for structural stability of systems with non-intersecting

static feedbacks*. An example of such a system is shown in Fig. 134.

Let us consider any system which differs from a single-loop system by the presence of any number of static non-intersecting feedbacks. We remove from the system all the feedbacks which include one or two stages of the first order, together with these included stages, replacing them by equivalent stages.** We can then find two stages, the jth and the kth, which satisfy the following two conditions:

(a) Feedback is transferred from the kth to the jth stage;

(b) The stages included by this feedback have no other feedback.

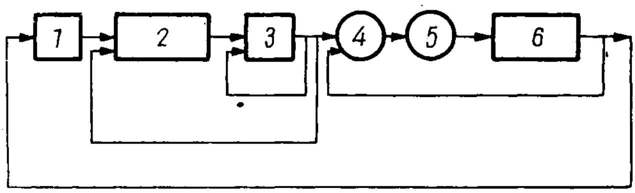


FIG. 134

Let us "excise" from the system this feedback and the stages it includes representing them separately, while instead of the "excised" stages we include in the system a chain of n single capacitance stages, where n is the degree of the characteristic equation of the "excised" part of the system. Again we find two stages, satisfying the above

* i.e. containing connexions such that the arrows which represent them on the structural diagrams do not intersect.

** Thus, for example, an astatic stage with feedback has the equation

$$\frac{dx_{out}}{dt} = k_1 (x_{in} - \varrho x_{out}) \quad \text{or} \quad T\frac{dx_{out}}{dt} + x_{out} = kx_{in},$$

where

$$T = \frac{1}{k_1 \varrho} \text{ and } k = \frac{1}{\varrho},$$

and therefore is equivalent to a single-capacitance stage.

Two astatic stages with static feedback have the equations

$$\frac{dx_1}{dt} = k_1 (x_{in} - \varrho x_2), \quad \frac{dx_2}{dt} = k_2 x_1 \quad \text{or} \quad \frac{d^2 x_2}{dt^2} + k_1 k_2 \varrho_2 = k_1 \cdot k_2 x_{in}.$$

Introducing new notation we obtain

$$\frac{T^2 d^2 x_{out}}{dt^2} + x_{out} = kx_{in} \quad \text{where } x_{out} = x_2;\ T^2 = \frac{1}{k_1 k_2 \varrho};\ k = \frac{1}{\varrho},$$

and consequently the two astatic stages with feedback are equivalent to one conservative stage, and so on.

condition, "excise" them from the system and similarly include a corresponding number of single capacitance stages in their place, and so on until the system becomes a single-loopsystem.

If the system contained h feedbacks and if only S of them may be excluded at once by introducing equivalent stages, then as a result we obtain $h - S + 1$ single-loop systems.

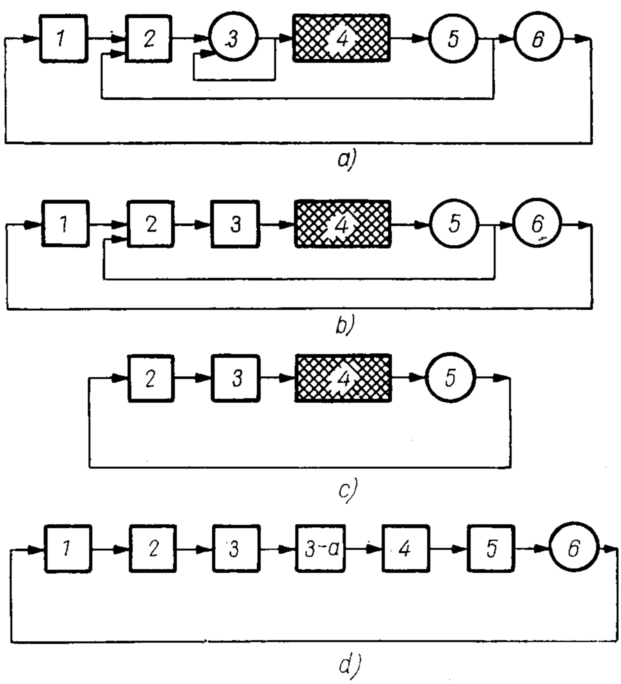


FIG. 135

We call this process "the decomposition of a multiple-loop system into single-loop systems". The value of this procedure is shown by the following theorem:

In order that a multiple-loop system with static non-intersecting feedbacks shall be structurally-stable it is sufficient (but not necessary) that this property be possessed by all the single-loop systems obtained by decomposition of the given multiple-loop system.

Let us find, by way of example, whether the system whose structural scheme is shown in Fig. 135*a*, is structurally stable.

The third stage together with its shunted feedback can be replaced by a single-capacitance stage without feedback which is equivalent to it. The resulting system is shown in Fig. 135*b*. It is completely equivalent to the original scheme, shown in Fig. 135*a*.

We now single out the second, third, fourth and fifth stages, together with their shunted feedbacks. The characteristic equation of the singled out system is of the fifth degree. We replace these stages by five single-capacitance stages (Fig. 135*c*). Thus the multiple-loop system (Fig. 135 *b* and *c*) can be broken down into two single-loops, both of which are structurally stable. Hence, the original multiple-loop system is also structurally stable.

Let us now find whether the system shown in Fig. 136 is structurally-stable when the engine possesses negative inherent regulation, and when there is friction in the sensor.

The structural scheme of the system is shown in Fig. 137*a*. It is decomposed into two single-loop systems (Fig. 137*b* and *c*). They are both structurally stable. Hence the original multiple-loop system of Fig. 136 is structurally stable.

(b) *The critical coefficient of amplification. The critical coefficient of amplification of a single-loop system without derivative action*

If a single-loop system consists only of single-capacitance and oscillatory stages, then it can always be made stable by the choice of its parameters ; to do this it is sufficient to choose the coefficient of amplification of the system, K, to be less than some critical value.

In fact, in this case the characteristic equation of the system has the form

$$\prod_{j=1}^{j=n} d_j(p) + K = 0.$$

where $K = \prod_{j=1}^{j=n} k_j$ is positive, and

$$d_j(p) = T_j p + 1$$

or

$$d_j(p) = T_j^2 p^2 + T_{jk} p + 1.$$

The hodograph $\prod_{j=1}^{j=n} d_j(i\,\omega) + K$ is obtained from the hodograph $\prod_{j=1}^{j=n} d_j(i\,\omega)$, by displacing it to the right, parallel to itself, by the amount K.

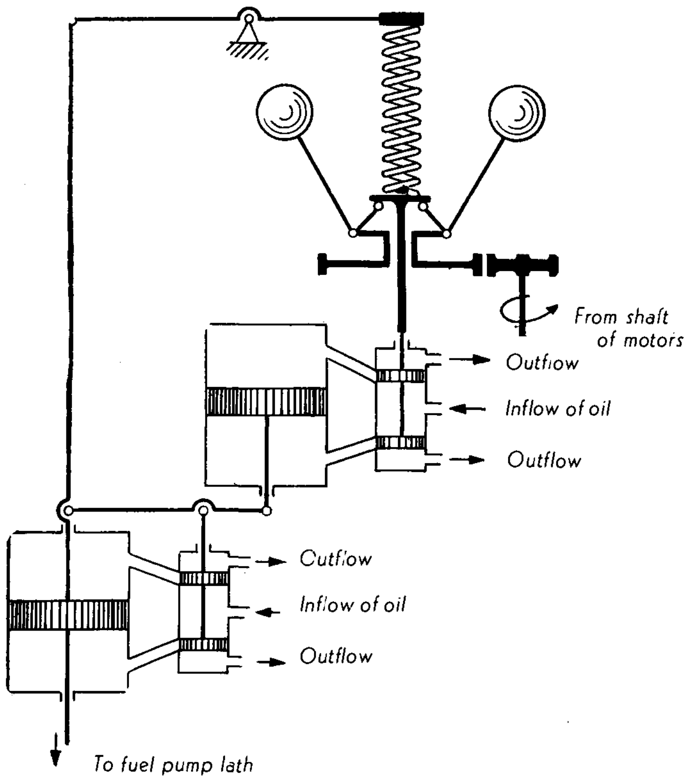


FIG. 136

The hodograph of the product $\prod_{j=1}^{j=n} d_j(i\,\omega)$ always satisfies the condition of the Mikhailov criterion. Suppose that its intersection with the negative u-axis nearest the origin of coordinates is the point u_0. Then for any $K > u_0$ the hodograph

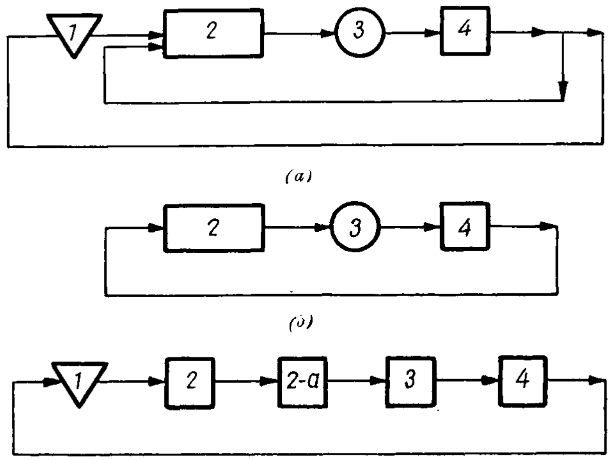
$$\prod_{j=1}^{j=n} d_j(i\omega) + K$$

also satisfies the conditions of the Mikhailov criterion and the system is stable.

The value $K = u_0$ is called the critical value and is denoted by K_{cr}. *The system is stable if* $K < K_{cr}$.

If

$$D(i\omega) = \prod_{j=1}^{j=n} d_j(i\omega) = u(\omega) + iv(\omega),$$



(a)

(b)

(c)

FIG. 137

then the critical value of the coefficient of amplification is determined from the condition

$$u(\omega_{\min}) = K_{cr},$$

where $\omega_{\min}$ is the least positive real root of the equation $v(\omega) = 0$, since the value of K_{cr} is equal to the section of the u-axis from the origin to the nearest point of intersection of this axis and the hodograph $D(i\omega)$.

The value of the coefficient of amplification determines the static error (see Chapter II, Section 7) ; the larger the coefficient of amplification, the smaller the non-conformity and the higher the static accuracy of the system.

Thus, *in every single-loop system the coefficients of amplification has a limiting value which ensures the stability of the control : in a static*

*system this value is determined by the minimal attainable static error in the system**.

From the point of view of the stability of the system it is desirable to have the smallest value of the coefficient of amplification. At the same time, from the point of view of the static accuracy of control it is desirable to increase the coefficient of amplification; it is possible to do this without destroying the stability only if K_{cr} is sufficiently large.

The critical value of the coefficient of amplification depends on the values of the time constants in the separate stages.

Let us confine ourselves to a system consisting of n single-capacitance stages. The characteristic equation of such a system is of the form

$$\prod_{j=1}^{j=n} (T_j p + 1) + K = 0.$$

Let us assume that all the time constants are equal, i.e.

$$T_1 = T_2 = \ldots = T_n.$$

In this case all the stages will have an identical hodograph.

Let the hodograph of the system intersect the u-axis at the value $\omega = \bar{\omega}$ and let R be the modulus, and a the argument of the radius-vector to the point of the hodograph of the stage corresponding to $\bar{\omega}$ (Fig. 138). If the system consists of n identical stages, then

$$a = \frac{\pi}{n} \quad \text{and} \quad K_{cr} = R^n.$$

But, from the triangle Oac (Fig. 138)

$$R = \sqrt{(T\bar{\omega})^2 + 1} \quad \text{and} \quad T\bar{\omega} = \tan\frac{\pi}{n}.$$

* Of course, only if an increase in static accuracy is obtained when K is increased for invariable values of all the time constants. We can often increase the static accuracy by changing those parameters determining K and T, and as K increases, K_{cr} is increased. A typical example of such a parameter is the rigidity of the piston of a sensor. With a decrease in rigidity, the static error is decreased. Then the coefficient of amplification of the sensor and its time constant increase simultaneously and, of course, so does K_{cr}.

Therefore the critical coefficient of amplification is

$$K_{cr} = R^n = [\sqrt{(T\bar{\omega})^2 + 1}]^n = \left[\sqrt{\tan^2 \frac{\pi}{n} + 1}\right]^n$$

and hence

$$K_{cr} = \frac{1}{\left[\cos\left(\frac{\pi}{n}\right)\right]^n}.$$

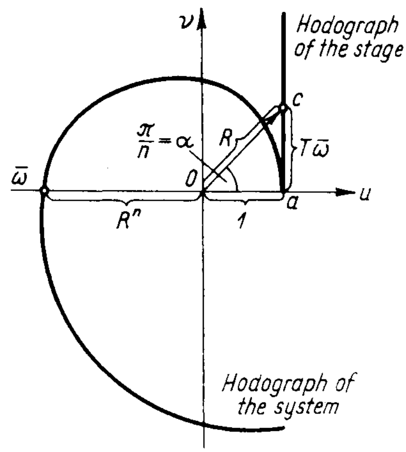


Fig. 138

Values of the critical coefficient of amplification calculated from this formula are set out in Table XII.

In present-day static control assemblies, in order for the system to be statically exact, it is often necessary that k should be from 50 to 100, or in some cases from 500 to 1,000.

Table XII

n	2	3	4	5	6
K_{cr}	∞	8	4	2·9	2·4

From the calculations made it follows that a system consisting of stages having identical (or, of course, approximately equal) time constants, is not suitable if the number of stages is more than two.

In order to ensure high values of K_{cr}, it is necessary to "move" the values of the time constants apart, to increase the range between them.

Let us now consider the case when the time constants form a geometric progression

$$\frac{T_1}{T_2} = \frac{T_2}{T_3} = \ldots = \frac{T_k}{T_{k+1}} = \ldots = \frac{T_{n-1}}{T_n} = \lambda .$$

The values of the critical coefficient of amplification in this case are set out in Table 13 for various values of λ and n.

TABLE XIII

	Number of stages			
	3	4	5	6
1	8	4	3	2·3
5	37	30	29	28
10	122	110	110	110
100	10,200	11,000	10,098	10,097

From Table XIII it follows that in order to obtain $K_{cr} > 50$ with a normal number of stages ($n < 7$) the value of λ must be of the order of 10.

Frequently static systems contain oscillatory stages in addition to single-capacitance stages, although the simple relations show how important it is to increase the ratio of the greatest time constant to the least, to "move" the values of the time constants apart from one another.

The increase in the critical coefficient of amplification due to the introduction of derivative action. Let us suppose that in the absence of derivative action the conditions for structural stability are satisfied and that derivative action is introduced only to increase the critical coefficient of amplification.

Let us number the stages in such a way that the first is the stage to which the derivative action is applied.

Then with first derivative action the equations of the control process after the application of the Laplace transform take the form

$$\left.\begin{aligned} d_1(p)\, L[x_1] &= -\, k_1 (1 + \varrho p)\, L[x_n]\,, \\ d_j(p)\, L[x_j] &= k_j\, L[x_{j-1}],\ j = 1, 2, \ldots, n, \end{aligned}\right\} \qquad (3.36)$$

where, as before, ϱ is different from zero if there exists first derivative action in the system. The characteristic equation of the system (3.36) reduces to the form

$$D(p) + K + K\,\varrho\,p = 0 \tag{3.37}$$

or

$$D(p) + K + Rp = 0, \tag{3.38}$$

where

$$R = K\,\varrho.$$

To find the conditions under which the introduction of derivative action increases the critical value of the coefficient of amplification, it is most convenient to construct the region of stability in the plane of the parameters K and ϱ. With this aim we first construct the region of stability in the plane of K and R and then pass into the plane of K and ϱ.

Let $D(i\,\omega) = u\,(\omega) + i\,v\,(\omega)$.

Putting $p = i\,\omega$ in (3.38) and equating the real and imaginary parts separately to zero, we obtain :

$$\left.\begin{aligned} K + u\,(\omega) &= 0\,, \\ R\omega + v\,(\omega) &= 0\,, \end{aligned}\right\} \tag{3.39}$$

giving

$$K = -\,u\,(\omega)\,;\; R = -\,\frac{v\,(\omega)}{\omega}\,. \tag{3.40}$$

The determinant of the system (3.39) is

$$\Delta = \begin{vmatrix} 1 & 0 \\ 0 & \omega \end{vmatrix} = \omega\,,$$

and therefore in the K, R-plane there is only one singular straight line, that corresponding to $\omega = 0$. Knowing the shape of the curve $D(i\,\omega)$ it is not difficult, using (3.40), to construct the D-partition boundary in the K, R-plane (Fig. 139). For this we need only change the signs of the abscissae and the ordinates and divide the latter by ω. The ordinate of the D-partition boundary then becomes zero for the same values of ω for which the ordinate of the hodograph $D\,(i\,\omega)$ becomes zero.

It is now not difficult to obtain the D-partition of the plane K, ϱ if we bear in mind that $\varrho = \dfrac{R}{K}$

The ordinate of the D-partition boundary in the K, ϱ-plane is defined as the tangent of the angle of inclination of the straight line joining the origin in the K, R-plane to the point of the D-partition boundary in this plane.

As an example, in Fig. 140d we complete the construction of the D-partition of the half-plane $K > 0$ for the case when the hodograph $D\,(i\,\omega)$ is as shown in Fig. 140a.

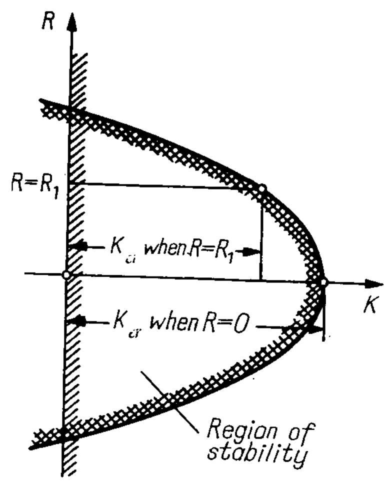


FIG. 139

In Fig. 140b we copy the hodograph of Fig. 140a changing only the signs of u and v. In Fig. 140c the ordinate of each point is divided by the value of ω corresponding to it. After shading the K, R-plane, the region of stability is seen. We are only interested in the half-plane $K > 0$. The previous construction is therefore carried out only for this half-plane (Fig. 140d) The abscissae of the curves in Fig. 140d and c coincide, and the ordinates in Fig. 140d are equal to $\tan\varphi$ (see Fig. 140c).

From Fig. 140d it follows that the greatest critical coefficient of amplification, $K_{\text{cr max}}$ is obtained when $\varrho = \varrho_a$. From the graph it is seen that $K_{\text{cr max}} = u_a$ and that the value of $K_{\text{cr max}} = u_a$ can be determined directly from the hodograph $D\,(i\,\omega)$, provided that (and

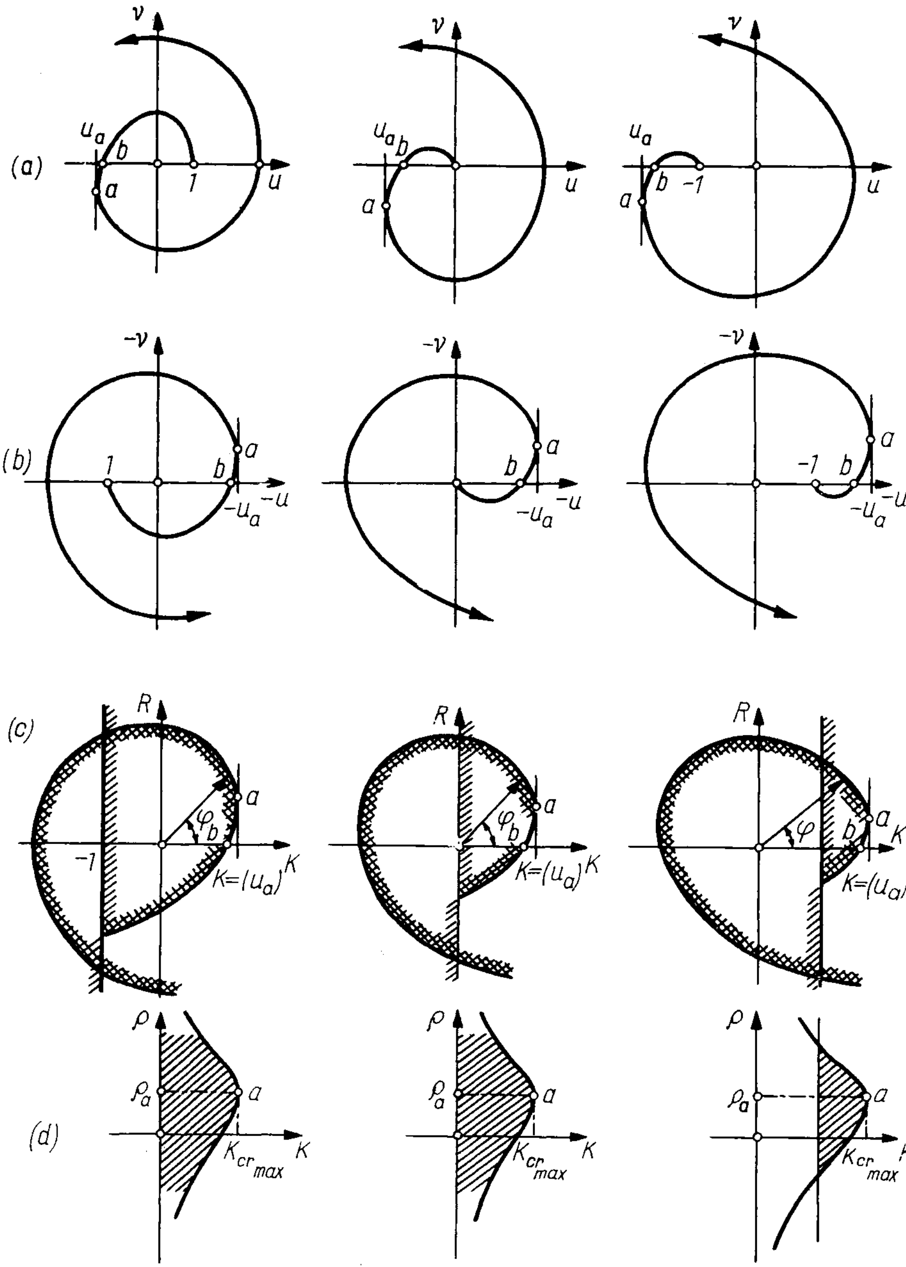


FIG. 140

this occurs most frequently) the hodograph $u + i\dfrac{v}{\omega}$ has no point of self-intersection. To do this we must draw the vertical tangent to the curve $D(i\,\omega)$ to the left of the v-axis and as near to it as possible. On the u-axis this tangent cuts off the value $u_a = K_{\text{cr max}}$ (Fig. 140a). If the point a lies in the third quadrant, then $\varrho_a > 0$ while if it lies in the second quadrant, then $\varrho_a < 0$.

The absolute value $|\varrho_a|$ is found from the equation

$$|\varrho_a| = \left|\frac{v_a}{\omega_a u_a}\right| \tag{3.41}$$

and can be calculated directly from the hodograph

Thus, *to determine the most advantageous coefficient for first derivative action from the point of view of an increase in the critical coefficient of amplification, it is necessary to construct the Mikhailov hodograph $D(i\,\omega)$ for an open system with derivative action and to find the abscissa u_a and the ordinate v_a of the point a where the tangent to the hodograph nearest the v-axis and lying to the left of it is vertical. The optimal value of ϱ_a is found from the equation*

$$|\varrho_a| = \left|\frac{v_a}{\omega_a u_a}\right|.$$

If the point a lies in the third quadrant, then $\varrho_a > 0$, i.e. in order to increase the critical coefficient of amplification we must introduce positive derivative action and the greatest coefficient of amplification which can be attained by introducing first derivative action will be $K_{\text{cr max}} = u_a$. If the point a lies in the second quadrant $\varrho_a < 0$, i.e. the critical coefficient of amplification is increased only if negative derivative action is introduced and then $K_{\text{cr max}} \leqslant u_a$.

This statement is only true, of course, if the hodograph $u + i\dfrac{v}{\omega}$ has no self-intersections. In the contrary case, the point a is not the point where the tangent is vertical, but is one of the self-intersecting points.

We recall that the modulus $d_j(i\,\omega)$ for astatic, single-capacitance and sufficiently well damped oscillatory stages increases monotonically as ω increases, and has no points of self-intersection. If the system

consists of such stages only, then the modulus $|D(i\omega)|$, equal to the product of the moduli $|d_j(i\omega)|$, also increases monotonically as ω increases. In this case the point a cannot lie in the second quadrant, and the critical coefficient of amplification can only increase when positive first derivative action is introduced if ϱ does not go outside the defined threshold. Here the introduction both of negative and of exceedingly strong positive ($\varrho > \varrho_a$) derivative action only lowers the critical coefficient of amplification.

In a system containing weakly damped oscillatory stages it is quite another matter. The modulus $|D(i\omega)|$ in such a system grows non-monotonically as ω increases, the hodograph $D(i\omega)$ can have points of self-intersection, and the point a can lie either in the third or in the second quadrant. In such cases positive derivative action can only decrease the critical coefficient of amplification, while negative derivative action can increase it, provided that

$$|\varrho| < |\varrho_a|.$$

The only exceptions to these rules are systems having characteristic equations of the second and third degrees, since a vertical tangent cannot be drawn to their hodograph $D(i\omega)$ in the finite part of the plane.

A system which has a second order characteristic equation and no derivative action is stable for any coefficient of amplification, and there is then no meaning in the problem of increasing the critical coefficient of amplification by introducing derivative action.

Hodographs $D(i\omega)$ of a system having a characteristic equation of the third order are shown in Fig. 141a. The construction of the region of stability in the K, R-plane and in the half-plane $K \geq 0$, ϱ is carried out in Fig. 141b and c.

In this case the critical coefficient of amplification increases monotonically as ϱ increases (for $\varrho > 0$), becoming infinite for the finite value $\varrho = \varrho_0$. If $\varrho > \varrho_0$ then the system is stable for any coefficient of amplification.

From this it follows that this remarkable property of single-loop systems is characteristic only of those systems with third degree characteristic equations.

For characteristic equations of higher degree ϱ becoming larger than the given threshold (the value ϱ_a) entails a risk of making a stable system unstable only because of the introduction of derivative action.

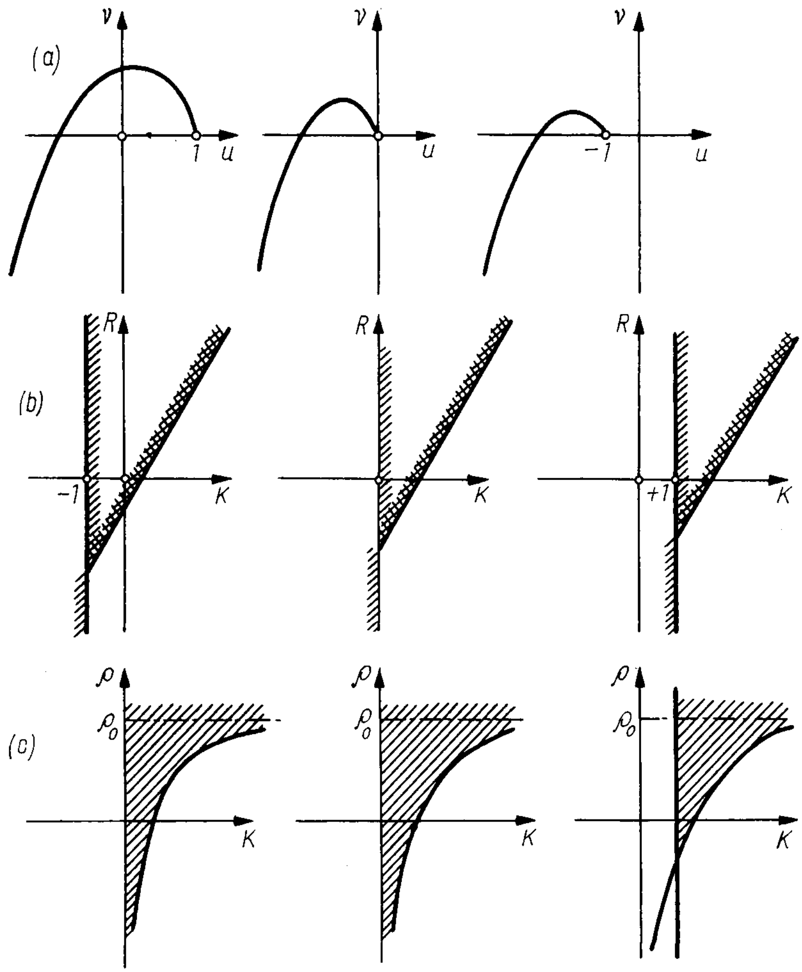


FIG. 141

EXAMPLE 1. Let us consider an indirect control assembly and let the time constant of the single-capacitance object be 5 sec, of the static servomotor be 1 sec, of the oscillatory sensor be $\sqrt{0{\cdot}1}$ sec, and of damping be 1 sec. Then the characteristic equation of the control process will be

$$D(p) + K = 0\,,$$

where

$$D(p) = (5p + 1)\,(0{\cdot}1p^2 + p + 1)\,(p + 1)\,,$$

and where only the value of K depends on the transmission ratios of the connexions.

The Mikhailov hodograph of the open system is

$$D(i\,\omega) = (5i\,\omega + 1)\,[(1 - 0{\cdot}1\omega^2) + i\,\omega]\,(i\,\omega + 1)$$

shown in Fig. 142.

From Fig. 142 we see that the critical coefficient of amplification of the system is equal to 12·1, and therefore it is not possible to obtain stable operation in the given system by changing the transmission ratios of the connexions, if the static error is to be less than

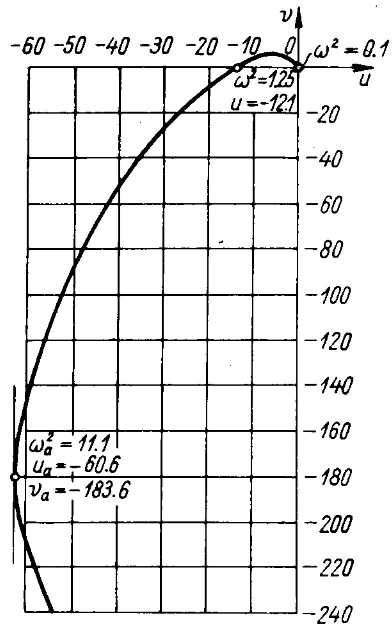


Fig. 142

$$\delta = \frac{1}{12{\cdot}1 + 1} \cong 0{\cdot}776,$$

i.e. less than 7·6%.

Now let derivative action be introduced into the system in order to decrease the static error.

To determine the maximum coefficient of amplification attainable in presence of derivative action, we draw the vertical tangent to the hodograph, as in Fig. 142.

The point of contact has the coordinates $u_a = -60{\cdot}6$, $v_a = 183{\cdot}6$, and it corresponds to $\omega_a = 3{\cdot}33$. This point lies in the third quadrant. Consequently, the critical coefficient of amplification can be increased, by introducing positive derivative action, to the value $K_{\text{cr max}} = 60{\cdot}6$. This maximum coefficient of amplification is attained when

$$\varrho_a = \frac{v_a}{\omega_a\, u_a} = \frac{183{\cdot}6}{60{\cdot}6.3{\cdot}33} = 0{\cdot}91.$$

The static error when derivative action is present can be decreased to

$$\delta = \frac{1}{60{\cdot}6 + 1} \cong 0{\cdot}016,$$

without destroying the conditions of stability, i.e. to 1·6%.

EXAMPLE 2. We consider the previous example supposing that $T_k = 0{\cdot}01$, the values of all the other time constants remaining unchanged.

The hodograph

$$D(i\omega) = (5i\omega + 1)\,[(1 - 0{\cdot}1\omega^2) + 0{\cdot}01i\omega]\,(i\omega + 1)$$

is represented in Fig. 143.

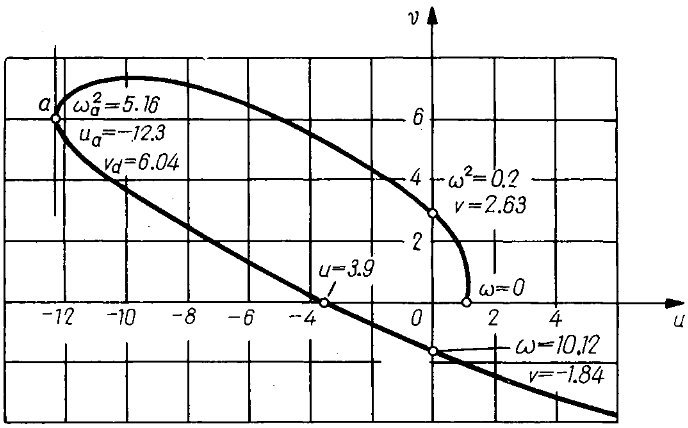


FIG. 143

In this case, when derivative action is absent, the critical coefficient of amplification is equal to only 3·9. The point where the tangent to the hodograph is vertical lies in the second quadrant. Its coordinates are $u_a = -12{\cdot}3$, $v_a = 6{\cdot}04$ and it corresponds to $\omega_a = 2{\cdot}27$. Consequently, for $T_k = 0{\cdot}01$, positive derivative action only reduces the coefficient of amplification.

In order to increase the coefficient of amplification it is necessary to introduce negative derivative action. Then the coefficient of amplification can only be increased up to $K_{\text{cr max}} = 12{\cdot}3$. This maximum coefficient of amplification is attained for

$$|\varrho_a| = \left|\frac{v_a}{\omega_a u_a}\right| = \frac{6{\cdot}04}{12{\cdot}3.2{\cdot}27} = 0{\cdot}216.$$

A similar investigation made for the case when second derivative action is introduced to increase the critical coefficient of amplification shows that the value of K_{cr} attained because of this action is unbounded

only if $n < 5$. For $n \geqslant 5$ the effect of the second derivative action is first to increase K_{cr} to some maximum value, and then reduce it to zero. Negative second derivative action only lowers K_{cr}.

4. Estimating the Stability of the Original System from the Stability of its Linear Model

In the previous sections we considered the question of the stability of the linear model of an automatic control system.

At the beginning of the chapter it was proved that at best the stability of the linear model indicates the stability of the given non-linear system for sufficiently small disturbances. In the present section this statement will be made more precise. Moreover, in some cases we can make a stronger assertion concerning the stability of the real system if it can be established that its linear model is stable.

To exemplify this, let us make the concept of stability which we introduced at the beginning of Chapter III more precise.

We call a system which contains non-linear elements *"slightly" stable* if there exists a region, however small, of initial deviations, such that as a result of a deviation which belongs to it, the steady conditions are restored (after a finite time or in the limit as $t \to \infty$).

Thus, *by saying that the controlled conditions are "slightly" stable we are only asserting the presence of a region of initial deviations with respect to which the system is stable (i.e. the presence of a region of stability) but are not defining any of its boundaries.*

Of course, the "slight" stability of a system does not prevent the system itself, for real initial deviations, behaving as if it were unstable, since the concept of slight stability does not take into account the fact that the region of stability of the system can be bounded. To speak of the stability of the real system it is necessary to compare the region of stability and the region of initial deviations which are possible under testing conditions.

We agree to say that the system is *"largely" stable* when the boundary of the region of initial deviations following which the controlled conditions are restored is defined and when it is shown that this region contains real initial deviations.

Finally, we agree to say that the system is *"decremented"* or *"unboundedly" stable ("wholly" stable)* when the region of initial

deviations, which lead to a restoration of the position of equilibrium, is not bounded at all. In this case the initial system possesses the same properties as its linear model ; from the fact that it is wholly stable for any initial deviation.

We now restrict ourselves to the case of a system which differs from a linear system by the presence of one non-linearity. In the general case the process in such a system is described by the equations:

$$\left.\begin{aligned} \dot{x}_1 &= \sum_{j=1}^{n} a_{1j} x_j + f(x_k), \\ \dot{x}_i &= \sum_{j=1}^{n} a_{ij} x_j, \quad i = 2, 3, \ldots, n, \end{aligned}\right\} \tag{3.42}$$

differing from the linear system

$$\left.\begin{aligned} \dot{x}_1 &= \sum_{j=1}^{n} a_{1j} x_j + a x_k, \\ \dot{x}_i &= \sum_{j=1}^{n} a_{ij} x_j, \quad i = 2, 3, \ldots, n, \end{aligned}\right\} \tag{3.43}$$

only by the presence of one non-linear function $f(x_k)$, in the first equation in place of ax_k.

Let us now find the region of values of a for which the system (3.43) is stable. Let it be established, for example, that the system (3.43) is stable for

$$a^* < a < a^{**}$$

and unstable for

$$a < a^* - \varepsilon \text{ and } a < a^{**} + \varepsilon,$$

at least for a sufficiently small positive number ε.

In other words, the values $a = a^*$ and $a = a^{**}$ are the boundary of the region of stability for a.

Further, in the construction of the linear model of the original system (3.42) let the non-linear function $f(x_k)$ be replaced by the linear function $a_0 x_k$, a_0 being chosen so that

$$a^* < a_0 < a^{**}. \tag{3.44}$$

It is immaterial how a_0 was determined, whether by a transition to small oscillation, i. e. by substitution of the $f = f(x_k)$ by the tangent at the point $x_k = 0$ or by experimental methods, i.e. by replacing this curve by a straight line, which, although it also passes through the point, is not coincident with the tangent. If the inequality (3.44) is satisfied, then the constructed linear model is stable. Can we then make the same conclusion concerning the stability of the original system (3.42)?

Let us consider two numbers a_1 and a_2 satisfying the inequality

$$a^* < a_1 < a_0 < a_2 < a^{**}. \tag{3.45}$$

Let us now construct in the f, x_k-plane two rays $f = a_1 x_k$ and $f = a_2 x_k$ and compare them with the curve $f = f(x_k)$.

There are three possible cases, represented in Fig. 144.

1. The curve $f = f(x_k)$ lies entirely, i.e. for all values of x_k possible when the system is tested, between the ray $f = a_1 x_k$ and the ray $f = a_2 x_k$ (Fig. 144*a*).

2. The curve $f = f(x_k)$ lies between the rays $f = a_1 x_k$ and $f = a_2 x_k$ only for sufficiently small x_k and intersects one of the rays for some value of x_k, say for $x_k = x_{k2}$ (Fig. 144*b*).

3. The curve does not lie between the rays for any x_k, however small (Fig. 144*c*).

Using the Lyapunov method, we establish that it is always possible, however the number a_0 was chosen when substituting the system (3.42) by its linear model (3.43) to find two numbers a_1 and a_2 such that the following assertions are true :

1. In the first case (Fig. 144*a*) the stability of the linear model indicates that the original system is unboundedly stable.

2. In the second case (Fig. 144*b*) the stability of the linear model shows only that the original system is "slightly" stable.

In addition, knowing the least value of x_k for which the curve intersects one of the rays, it is possible to determine the region belonging to the region of stability. Comparing it with the given region of initial deviations, it is sometimes possible to establish also that the stability of the linear model indicates that the system is "largely" stable.

3. In the third case (Fig. 144*c*) the stability of the linear model does not even indicate the slight stability of the original system.

If the linear model is constructed by means of a transition to small oscillations, i.e. if $f = a_0 x_k$ is the tangent to the curve $f = f(x_k)$ at the origin of coordinates, then from (3.45), the tangent lies between

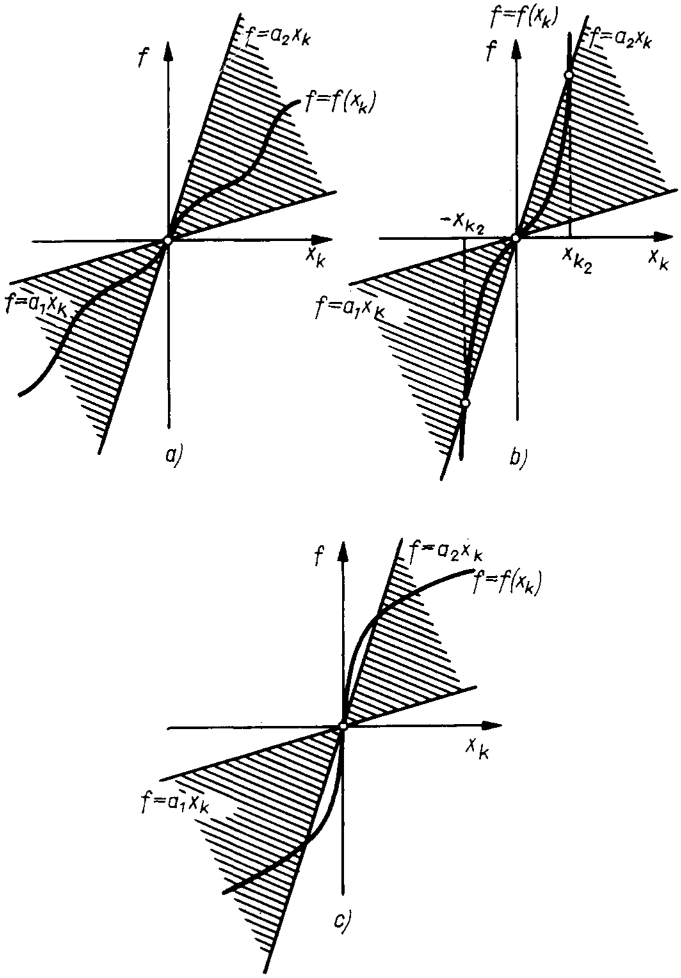


FIG. 144

the rays and the case of Fig. 144c is not possible. When the linear model has been constructed by transition to small oscillations the stability of the linear model always indicates the "slight" stability of the original system. We may in this case sometimes establish by the above

method that the original system is "largely" stable or even (as in Fig. 144*a*) that it is unboundedly stable.

If the linear model is constructed by the experimental averaging of non-linearities, for example, if experimentally found frequency characteristics are used or if the parameters of the elements (time constant, coefficients of amplification, etc.) are found from experimentally derived time characteristics, then the case of Fig. 144*c* is possible and the stability of the linear model does not ensure that the original system is even "slightly" stable.

All that has been said can be directly applied to systems which contain any number of linearizable non-linear functions of any number of arguments or which include parameters that are functions of time.

Returning to the case of the system (3.42), containing only one non-linear function, it is natural to ask :

Is it not possible to move the rays to the boundary of stability, i.e. is it not possible for the given deductions to remain in force if we replace the ray $f = a_1 x_k$ *by the ray* $f = a^* x_k$, *and the ray* $f = a_2 x_k$ *by the ray* $f = a^{**} x_k$?

So far we have not succeeded in finding a single example to refute this suggestion when some easily satisfied conditions are laid on the function $f(x)$. But it has been found possible to prove this assertion only for systems of the second order, or for separate special cases of systems of higher order, although the attention both of mathematicians and of engineers have been attracted to the problem.

When the initial system contains non-linearizable non-linearities, the situation is considerably more complicated. Up to now we have not succeeded in finding general methods enabling us to construct a linear model in these cases too, in order to answer the question of whether the stability of the linear model indicates the stability of the original system, even if only the "slight" stability. We have succeeded in solving only a number of special non-linear problems with respect to non-linearizable non-linearities. These solutions show that sometimes the presence of non-linearizable non-linearities leads to the fact that it is not possible to judge the stability of the system even for "slight" stability from the stability of its linear model, if the non-linearizable non-linearities were rejected during its construction (for example, dry friction in the sensor of an indirect action controller). However, cases where the non-linearizable non-linearities tend to favour the stability of the system are also known.

5. Concluding Remarks

When the linear model of the given system is given by its equation of motion (transfer function), we judge the stability of the linear model from its characteristic equation. This equation is obtained by equating the denominator of the transfer function to zero. To judge the stability it is most convenient to use a reduction which successively lowers the degree of the equation (p. 181), or to reduce the Hurwitz determinant to diagonal form (p. 176 – 179). The region of stability for one or two parameters which enter the characteristic equation linearly can be found by the relatively simple construction of the D-partition boundary.

When the properties of the linear model of the system are given by its amplitude-phase characteristic (or its amplitude and phase logarithmic characteristics), an estimate of the stability of the linear model can be made directly from the graph of these characteristics. But then in the case of non-linear, linearizable systems the question about the validity of such judgements even for small disturbances remains open. The results of such an analysis are unconditionally valid only when the original real system is linear for all possible ranges of change in all its generalized coordinates.

If the investigation results in establishing the instability of the linear model, then a number of qualitive considerations about methods of ensuring stability can be made without further calculation. For this we use theorems concerning structural stability, or concerning the critical coefficient of amplification and methods of increasing it. In particular, in a structurally-unstable system to ensure stability requires, above all, a change in the structural scheme. If the scheme is structurally stable, but for the given values of the parameters stability does not exist, then by a sufficient decrease in the coefficient of amplification stability can always be secured. In doing this, in order not to increase the static error of the system (if the system is static) past its admissible value, it is necessary to increase the range between the extreme values of the time constants. To this end derivative action is unconditionally effective for systems of the third order (with first derivative action) and for systems up to and including the fourth order (for second derivative action). For a higher order system the use of derivative action requires great care: excessive (or sometimes any) derivative action can only interfere with the stabilization of the system and worsen the conditions of its stability.

In concluding the analysis of stability, it is necessary to remember that the stability of the linear model of the system is verified, but by no means that of the actual system. Only experiment and experience of operation prove whether the computational linear model was reasonably chosen, and whether a linear analysis is generally suitable for the investigation of the stability of the system. Only the accumulated experience of calculation of systems of a defined type enables us to choose the computational model with confidence, to estimate beforehand what we may or may not ignore in its formation.

CHAPTER IV

THE CONSTRUCTION AND EVALUATION OF THE PROCESSES IN THE LINEAR MODEL OF A SYSTEM OF AUTOMATIC CONTROL

1. General Considerations

Until now we have only considered the conditions of stability for the linear model of a control system, and not the character of the control process. Obviously, the presence of stability is not a sufficient condition for the normal operation of the system since, for example, the damping of the oscillations may be too slow or the deviations in the controlled coordinate during the control process may exceed the allowable limits, and so on.

In automatic control assemblies great importance is attached to the study of the transient process during the transition from one set of conditions to another (for example, when the load is altered). It is then no longer possible to regard the deviations as small, and the investigation of the transient process requires an analysis of the initial non-linear equations in considerably greater measure than does the investigation of stability. The construction of the process in the linear model enables us to evaluate the process in the real assembly only for small disturbances, and also only if all its elements are linearizable.

For this reason, we give below only a short account of some questions connected with the construction and evaluation of the quality of the control process in the linear models of control systems.

The control process is called into action by external disturbances applied to the system. In order to construct the control process or to evaluate it, it is first of all necessary to elucidate the character of the external actions causing the process.

(a) *External actions*

There are three forms of external disturbance, depending on what external action is stipulated, namely : load, tuning, or noise.

The *load* is an external action applied to the controlled object, independently of the controller and stipulated by a change in the operating conditions of the controlled object.

The tuning includes those disturbances which are applied deliberately to any element of the controller with the purpose of changing the value of the controlled coordinate maintained by the controller.

Noise consists of those external actions on the separate elements of the controller or controlled object which are not necessary for normal operation of the assembly but which exist only because they cannot be removed from its construction.

Thus, for example, in the variable speed controller of a transport diesel, assembled in an automobile, the change in tractive force transmitted to the engine is the load, the action of the driver's foot on the accelerator transmitted to the controller is the tuning, and the action transmitted to the controller as a result of shaking, vibrations of the engine, bumps in the road and so on, is the noise.

In an electronic controller, controlling the supply of a steam-boiler, the actions arising as a result of a change in the selection of steam by the user is the load, the action on the setter of the controller made in order to change the maintained level is the tuning, and the actions arising as a result of noise in the amplifier valves constitute the noise.

In most cases the external disturbance of any of these three forms is a complicated function of time, and in control theory we are very rarely given the task of determining the reaction of the system to external actions given by functions corresponding exactly to the real disturbances acting on the system. If the disturbances were known beforehand, the problem of control would be greatly simplified and would be replaced by the problem of compensation of the disturbance. But usually the external actions cannot be exactly defined beforehand, and the problem is simplified by the idealization of the disturbance, i.e. by replacing the real disturbance by a simpler, typical, function of time.

The most commonly used if such typical functions of time are the unit function* $\mathbf{1}$, the product of the unit function and a sinusoid, $\mathbf{1}\, A \sin \omega t$, and the product of the unit function and the exponential $\mathbf{1}\, Ae^{rt}$.

The function $f(t) = \mathbf{1}\, A \sin \omega t$ is used when the external action on the system is periodic.** The steady oscillations of any coordinate (including the controlled coordinate) are completely defined by the amplitude-phase characteristic of this closed system.

The function $f(t) = \mathbf{1} \cdot A$ is used when the process is caused by an increase or decrease in the load on the controlled object, by a rapid displacement of the controller setting mechanism, and so on. Control processes caused by external actions of this kind are called transient***. For linear systems, the processes caused by an action of the form $f(t) = \mathbf{1} \cdot A$ differ from processes caused by the action $f(t) = \mathbf{1}$ only by the scale along the y-axis being increased A times. Hence, in the construction and evaluation of transient processes we can restrict ourselves to the consideration of unit action $f(t) = \mathbf{1}$.

The main content of this chapter consists of methods for the construction and evaluation of transient processes.

The function

$$f(t) = \mathbf{1} \cdot A(1 - e^{-rt}), \tag{4.1}$$

where r is positive and A is any constant quantity, is used when the external action builds up smoothly. As an example, we may take the slow readjustement of the controller, the gradual removal of the load from the object, and so on.

If we know the transient process, it is easy to construct the process caused by the action (4.1).

* Functions $\mathbf{1}$ such that $\mathbf{1} = \begin{cases} 0 \text{ for } t < 0, \\ 1 \text{ for } t \geq 0 \end{cases}$ are called unit functions. The product $\mathbf{1} \cdot f(t)$, used below, is the function which is equal to zero for $t < 0$ and equal to $f(t)$ for $t > 0$. If the disturbance is expressed as such a product, then this means that it does not exist until $t = 0$, that it first acts on the system at the moment $t = 0$, and that, later, it is expressed by the function $f(t)$. In the theory of servomechanisms, in addition to the functions given in the text, the products of the unit function with a linear function and with various parabolas (i. e. $\mathit{1}kt^r$ where $r = 1, 2, \ldots$) are also used.

** For example, several forms of noise, including vibrations transmitted to the body of the instrument.

*** If we know the course of the transient process defined in this way, we can construct the process in the same system for any other action.

In some cases, in particular when calculating the influence of noise, the substitution for the true disturbance by some typical function is of no use. Noise, and sometimes loads too, change discontinuously and in a random manner. It is expedient to evaluate such disturbances by using their static characteristics. Of course, as a result of the calculation in these cases we obtain only the static characteristics of the process caused by these disturbances, for example, the mean square deviation of the controlled parameter (see Section 8 of this chapter).

(b) *Parameters characterizing the quality of the transient process*

Usually the following restrictions are placed on the transient process :

1. The transient process must be completed in a certain time t_p, called the settling time. In theory, the transient process in linear systems is continued for an unlimited period. In practice, however, the transient process is completed as soon as the deviation in the controlled parameter does not exceed some defined limit.

In the case of static systems we often consider that the transient process is completed at the instant of time at which the value of the deviation of the coordinate differs from the steady value by not more than 5 per cent of the static error.

For astatic systems we regard the transient process as completed when the coordinate value does not exceed a definite part of its normal value, the fractions indicated often being considerably less than 5 per cent.

2. The greatest deviation in the controlled coordinate of the control process from the value which must be set up after the completion of the transient process must not exceed an allowable quantity.

In static systems, the greatest deviation of the controlled coordinate during the transient process, of the same sign as the static deviation, is sometimes called the overshoot. Usually not only overshoot but also the total greatest deviation in the controlled coordinate whatever its sign is important.

Overshoot is often estimated in percentages of the static error caused by the same disturbance.

For astatic systems the term overshoot formulated in this way is inapplicable. In this case we speak of the largest dynamic deviation, meaning by this the greatest deviation of the coordinate from the standard value (in percentages) during the transient process.

Figure 145 shows the boundary of the region inside which the transient process can lie when it satisfies the above two requirements.

Sometimes additional conditions are laid on the course of the transient process : for example, we can require that the process be monotonic or that the number of oscillations during the process be not greater than a given number, and so on.

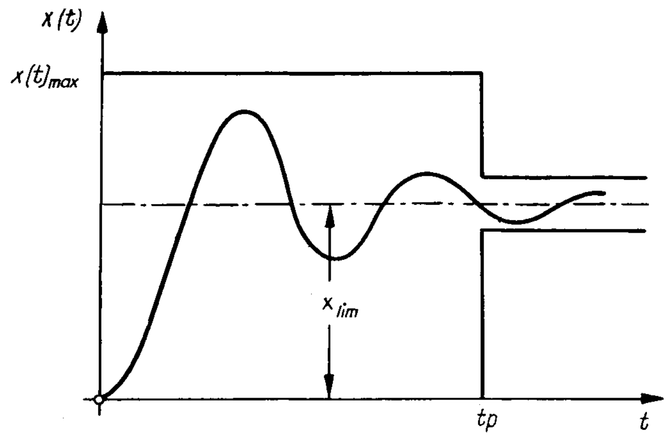


FIG. 145

If methods existed for the exact estimation of the above basic parameters, characterizing the quality of the process from the form of the differential equations, then there would be no need to carry out completely the integration of these equations. Such methods have not yet been devised, and although it is often not important to know the shape of all the integral curves, but only some parameters characterizing their shape, the only completely reliable method at present for calculating these parameters consists in the construction of the integral curves.

To construct the curve of the control process for some fixed values of the parameters, it is necessary to perform labour-consuming calculations. To select the optimal values of the parameters these calculations must be repeated many times. As a result a particular

importance is attached to various approximate and indirect methods of evaluating the process, which do not require the construction of the integral curves.

In proceeding to an account of the methods for constructing and evaluating the transient process, we will assume everywhere in this chapter that the given system is stable.

2. The Construction of the Process from the Transfer Function of the System

(a) *Description of the method*

For any automatic control system the Laplace transform of one of the system coordinates can be written in the following form* :

$$L[x(t)] = W(p)\,L[f(t)] = \frac{A_1(p)}{B_1(p)}\,\frac{S(p)}{R(p)} = \frac{A(p)}{B(p)}, \tag{4.2}$$

where $W(p) = \dfrac{A_1(p)}{B_1(p)}$ is the transfer function of the system (from the given point of application of the external action to the coordinate x we are considering), and $L[f(t)] = \dfrac{S(p)}{R(p)}$ is the Laplace transform of the external action (for the typical actions listed above $\dfrac{S(p)}{R(p)}$ is a rational, fractional function of p).

The change as a function of time of the coordinate $x(t)$ caused by an action $f(t)$, from the second theorem of the Heaviside expansion, is defined by the function*

$$x = \sum_{k=1}^{n} \frac{A(p_k)}{B'(p_k)}\,e^{p_k t}, \tag{4.3}$$

where p_k is a root of the equation $B(p) = 0$ (we assume that there are no repeated roots).

When $f(t) = 1$, we have :

$$L[f(t)] = \frac{1}{p},\quad B(p) = pB_1(p),\, A(p) = A_1(p),$$

* See Appendix 1 (p. 489).

* See Appendix 1 (p. 489).

where $B_1(p)$ is the denominator of the transfer function $W(p)$. By putting $B_1(p)$ equal to zero, we obtain the characteristic equation of the system. In this particular case, instead of (4.3), we obtain

$$x = \frac{A_1(0)}{B_1(0)} + \sum_{k=1}^{n} \frac{A_1(p_k)}{p_k B_1'(p_k)} e^{p_k t}. \tag{4.4}$$

The term $\dfrac{A_1(0)}{B_1(0)}$ determines the steady deviation ; each term corresponding to the real root of the characteristic equation gives an exponential function of the form $x_j = A_j e^{p_j t}$, and each pair of terms corresponding to a pair of complex conjugate roots $p_k, p_{k+1} = a_k \pm i\beta_k$ gives a function of the form

$$x_j = A_j e^{at} \sin(\beta t + \varphi_j).$$

Formulae for the calculation of A_j and φ_j are given in the Appendix.**

The problem of the construction of the transient process is then reduced to the calculation, first, of all the roots of the characteristic equation, and then of all the coefficients A_j and φ_j. It is then necessary to construct the exponential functions, the product of the exponential and trigonometric functions and to sum the ordinates of the constructed graphs of these functions corresponding to the same values of t. The greatest difficulty here is caused by the approximate determination of the roots of the characteristic equation.

If the external action is different from $f(t) = 1$, the process tends to equation (4.3) for which it is necessary to know the roots of the equation

$$B(p) = B_1(p) R(p) = 0.$$

But for typical actions the roots of $R(p)$ can be determined at once, since $R(p)$ is a polynomial of the first or second degree. In this case the difficulty consists in the determination of the roots of the characteristic equation $B_1(p) = 0$. Of the many methods used for this purpose, we describe one iterative method.

** See p. 491.

(b) *An iterative method for the approximate calculation of roots of the characteristic equation*

We consider the characteristic equation

$$F(p) = p^n + a_1 p^{n-1} + a_2 p^{n-2} + \dots + u_{n-1} p + a_n = 0, \quad (4.5)$$

where $a_1, a_2, a_3, \dots, a_n$ are real numbers such that all the roots have a negative real part* (the system is stable).

It is required to find the approximate values of the real and complex roots of this equation. To do this, we divide equation (4.5) by the polynomial

$$F_1(p) = p^{n-1} + (a_1 - \alpha) p^{n-2} + (a_2 - \beta) p^{n-3} + \dots + (a_{n-1} - \omega),$$

where $\alpha, \beta, \dots, \omega$ are as yet unknown constants.

Having done this division, we obtain a quotient

$$p + \alpha$$

and a remainder

$$[\beta - \alpha(a_1 - \alpha)] p^{n-2} + \dots + [a_n - \alpha(a_{n-1} - \omega)].$$

We denote the remainder by $F_2(p)$. Then

$$F(p) = (p + \alpha) F_1(p) + F_2(p).$$

If the division results in no remainder $-\alpha$ is a root of the equation $F(p) = 0$.

If we make $F_2(p)$ identically equal to zero, and from this condition find $\alpha, \beta, \gamma, \dots, \omega$, then α will be the first root of the equation $F(p) = 0$. To find it, we equate to zero all the coefficients in $F_2(p)$ separately. As a result we obtain the system of equations

$$\left.\begin{aligned} \alpha &= \frac{a_n}{a_{n-1} - \omega} \\ \beta &= \alpha(a_1 - \alpha), \\ \gamma &= \alpha(a_2 - \beta), \\ \delta &= \alpha(a_3 - \gamma), \\ &\dots\dots\dots \\ \omega &= \alpha(a_{n-2} - \tau). \end{aligned}\right\} \quad (4.6)$$

* The method is correct if the multiplicity of the real roots in (4.5) is not greater than two, but there can also be several pairs of repeated complex roots. In practice, this restriction is unimportant, since by a small change in the coefficients (which are all determined equally approximately) we can avoid multiple roots.

The determination of the root α reduces to a simple iteration. Putting $\omega = 0$ we obtain from the above formulae :

$$\begin{aligned} \alpha_1 &= \frac{a_n}{a_{n-1}} \\ \beta_1 &= \alpha_1 (a_1 - \alpha_1) , \\ \gamma_1 &= \alpha_1 (a_2 - \beta_1) , \\ \delta_1 &= \alpha_1 (a_3 - \gamma_1) , \\ &\cdots\cdots \\ \omega_1 &= \alpha_1 (a_{n-2} - \tau_1) . \end{aligned}$$

We arrive at $\omega = \omega_1 \neq 0$. Hence, the arbitrary hypothesis that $\omega = 0$ was incorrect.

Putting now $\omega = \omega_1$ we obtain

$$\left.\begin{aligned} \alpha_2 &= \frac{a_n}{a_{n-1} - \omega_1} , \\ \beta_2 &= \alpha_2 (a_1 - \alpha_2) , \\ \gamma_2 &= \alpha_2 (a_2 - \beta_2) , \\ \delta_2 &= \alpha_2 (a_3 - \gamma_2) , \\ &\cdots\cdots \\ \omega_2 &= \alpha_2 (a_{n-2} - \tau_2) . \end{aligned}\right\} \tag{4.7}$$

If $\omega_2 = \omega_1$ then α_2 is a root of the given equatio. But if $\omega_2 \neq \omega_1$, then the iteration procedure continues.

EXAMPLE. We consider the equation

$$x^3 + 6{\cdot}32x^2 + 27{\cdot}5x + 31{\cdot}6 = 0 .$$

Let us put $a_1 = 6{\cdot}32$, $a_2 = 27{\cdot}5$, $a_3 = 31{\cdot}6$.

To determine the first root we make use of the formulae (4.6) and (4.7) given above. We rewrite them for an equation of the third degree:

$$\alpha = \frac{a_3}{a_2 - \beta} , \quad \beta = \alpha (a_1 - \alpha) .$$

Putting $\beta_0 = 0$ we obtain

$$\alpha_1 = \frac{31{\cdot}6}{27{\cdot}5} = 1{\cdot}15, \ \beta_1 = 1{\cdot}15 (6{\cdot}32 - 1{\cdot}15) = 5{\cdot}95 .$$

If we put $\beta_1 = 5{\cdot}95$, then

$$\alpha_2 = \frac{31{\cdot}6}{27{\cdot}5 - 5{\cdot}95} = 1{\cdot}47, \ \beta_2 = 1{\cdot}47 (6{\cdot}32 - 1{\cdot}47) = 7{\cdot}12.$$

If $\beta_2 = 7{\cdot}12$ we have

$$a_3 = \frac{31.6}{26{\cdot}5 - 7{\cdot}12} = 1.55, \ \beta_3 = 1{\cdot}55\,(6{\cdot}32 - 1.55) = 7.4.$$

If $\beta_3 = 7{\cdot}4$ we obtain

$$a_4 = \frac{31{\cdot}6}{27{\cdot}4 - 7{\cdot}4} = 1{\cdot}57, \ \beta_4 = 1{\cdot}57\,(6{\cdot}32 - 1{\cdot}57) = 7{\cdot}48.$$

Noting that $\beta_4 \approx \beta_3$ we stop the iteration process. Hence $a = -1{\cdot}57$ is the approximate value of the root of the given equation. In this case

$$F_1(p) = x^2 + (6{\cdot}32 - 1{\cdot}57)\,x + (27{\cdot}5 - 7{\cdot}48)\,,$$

i.e.

$$F_1(p) = x^2 + 4{\cdot}75x + 20{\cdot}02 = 0.$$

The roots of this quadratic equation are the other two roots of the original third degree equation.

If the iteration diverges, this means that the nearest root to the imaginary axis is complex, and the iteration must be done in another way. We must now divide $F(p)$ by a polynomial of degree two lower than itself (since we now determine two roots at once) :

$$F_3(p) = p^{n-2} + (a_1 - a)\,p^{n-3} + (a_2 - \beta)\,p^{n-4} + (a_3 - \gamma)\,p^{n-5} + \\ + \ldots + (a_{n-3} - \varrho)\,p + (a_{n-2} - \tau)\,.$$

Then

$$F(p) = \left\{p^2 + a\,p + [\beta - a\,(a_1 - a)]\right\} F_3(p) + F_4(p),$$

where the remainder after the division is

$$F_4(p) = \left\{\gamma - a\,(a_2 - \beta) - (a_1 - a)\,[\beta - a\,(a_1 - a)]\right\} p^{n-3} + \ldots \\ \ldots + a_n - (a_{n-2} - \tau)\,[\beta - a\,(a_1 - a)]\,.$$

If $F_4(p) = 0$, then we can factorize $F(p)$ into the factors $F_3(p)$ and $p^2 + a\,p + [\beta - a\,(a_1 - a)]$.

Solving the quadratic equation

$$p^2 + a\,p + [\beta - a\,(a_1 - a)] = 0,$$

we find the two complex roots of the equation $F(p) = 0$. From the condition $F_4(p) = 0$, equating all its coefficients to zero, we obtain a system of equations determining the quantities $\alpha, \beta, \gamma, \delta, \ldots, \tau$:

$$\left.\begin{aligned} \alpha &= \frac{a_{n-1}}{a_{n-2} - \tau} - \frac{a_n (a_{n-3} - \varrho)}{(a_{n-2} - \tau)^2}, \\ \beta &= \alpha (a_1 - \alpha) + \frac{a_n}{a_{n-2} - \tau}, \\ \gamma &= \alpha (a_2 - \beta) + \frac{a_n}{a_{n-2} - \tau} (a_1 - \alpha), \\ \delta &= \alpha (a_3 - \gamma) - \frac{a_n}{a_{n-2} - \tau} (a_2 - \beta). \end{aligned}\right\} \tag{4.8}$$

Now the iteration continues as follows :

Putting $\tau = 0$ and $\varrho = 0$ in the first equation, we find α ; putting this value of α and $\tau = 0$ in the second equation we find β and so on ; continuing this process we find the values of ϱ_1 and τ_1. Using these values we find the second approximation to the values of $\alpha, \beta, \gamma, \ldots, \tau$. This operation is continued until the iteration converges.

Having determined the values of $\alpha, \beta, \gamma, \ldots, \tau$ in this way, and having solved the quadratic equation

$$p^2 + \alpha p + [\beta - \alpha (a_1 - \alpha)] = 0,$$

we find the two complex roots of the equation $F(p) = 0$.

The remaining roots are found from the equation $F_3(p) = 0$ by a similar iteration process. If the distribution of the roots is not known, then we begin the process of iteration by applying the formulae for determining real roots.; if the process diverges, then we apply the formulae for determining complex roots.

Table 14 gives ready formulae for the determination by iteration of the roots of equations of up to and including the sixth degree. With the help of these formulae we can determine the smallest or largest (in modulus) root. Having found it we can reduce the degree of the equation and use the given formulae again.

The described iterative method is convenient in that the roots are determined in sequence from the smallest to the largest in modulus.

TABLE

THE DETERMINATION OF THE ROOTS OF

Form of Equation	Distribution of Roots
Third degree $x^3 + a_1 x^2 + a_2 x + a_3 = 0$	Root smallest in modulus is real
	Roots smallest in modulus are complex conjugates
Fourth degree $x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0$	Root smallest in modulus is real
	Roots smallest in modulus are complex conjugates or both real
	Root largest in modulus is real
Fifth degree $x^5 + a_1 x^4 + a_2 x^3 + a_3 x + a_5 = 0$	Root smallest in modulus is real
	Roots smallest in modulus are complex conjugates or both real

XIV

ALGEBRAIC EQUATIONS BY ITERATION

Working Formulae	Expansion in Factors
$\alpha = \dfrac{a_3}{a_2 - \beta}$; $\beta = \alpha(a_1 - \alpha)$	$(x + \alpha)\,[x^2 + (a_1 - \alpha)\,x + a_2 - \beta]$
$\alpha = \dfrac{a_2}{a_1 - \alpha} - \dfrac{a^3}{(a_1 - \alpha)^2}$	$(x + a_1 - \alpha)\left[x^2 + \alpha x + \dfrac{a_3}{a_1 - \alpha}\right]$
$\alpha = \dfrac{a_4}{a_3 - \gamma}$; $\beta = \alpha(a_1 - \alpha)$; $\gamma = \alpha(a_2 - \beta)$	$(x + \alpha)\,[x^3 + (a_1 - \alpha)\,x^2 +$ $+ (a_2 - \beta)\,x + a_3 - \gamma]$
$\alpha = \dfrac{a_3}{a_2 - \beta} - \dfrac{a_4(a_1 - \alpha)}{(a_2 - \beta)^2}$; $\beta = \alpha(a_1 - \alpha) + \dfrac{a_4}{a_2 - \beta}$	$[x^2 + (a_1 - \alpha)\,x + a_2 - \beta]\times$ $\times\left[x^2 + \alpha x + \dfrac{a_4}{a_2 - \beta}\right]$
$\alpha = \dfrac{a_2}{a_1 - \alpha} - \dfrac{a_3}{(a_1 - \alpha)^2} +$ $+ \dfrac{a_4}{(a_1 - \alpha)^3}$	$(x + a_1 - \alpha)\Big\{x^3 + \alpha x^2 +$ $+ [a_2 - (a_1 - \alpha)\,\alpha]x + \dfrac{a_4}{a_1 - \alpha}\Big\}$
$\alpha = \dfrac{a_5}{a_4 - \delta}$; $\beta = \alpha(a_1 - \alpha)$; $\gamma = \alpha(a_2 - \beta)$; $\delta = \alpha(a_3 - \gamma)$	$(x + \alpha)\,[x^4 +$ $+ (a_1 - \alpha)\,x^3 + (a_2 - \beta)\,x^2 +$ $+ (a_3 - \gamma)\,x + a_4 - \delta]$
$\alpha = \dfrac{a_4}{a_3 - \gamma} - \dfrac{a_5(a_2 - \beta)}{(a_3 - \gamma)^2}$; $\beta = \alpha(a_1 - \alpha) + \dfrac{a_5}{a_3 - \gamma}$; $\gamma = \alpha(a_2 - \beta) + \dfrac{a_5}{a_3 - \gamma}(a_1 - \alpha)$ or other formulae, obtained from the system of equations (4.6)	$\left(x^2 + \alpha x + \dfrac{a_5}{a_3 - \gamma}\right)\times$ $\times[x^3 + (a_1 - \alpha)\,x^2 + (a_2 - \beta)\,x +$ $+ a_3 - \gamma]$

TABLE XIV

THE DETERMINATION OF THE ROOTS OF

Form of Equation	Distribution of Roots
Fifth degree (contd.) $x^5 + a_1 x^4 + a_2 x^3 + a_3 x^2 + a_4 x + a_5 = 0$	Root largest in modulus is real
	Roots largest in modulus are complex conjugates or both real
Sixth degree $x^6 + a_1 x^5 + a_2 x^4 + a_3 x^3 + a_4 x^2 + a_5 x + a_6 = 0$	Root smallest in modulus is real
	Roots smallest in modulus are complex conjugates or both real

(contd.)

ALGEBRAIC EQUATIONS BY ITERATION

Working Formulae	Expansion in Factors
$\begin{aligned} \alpha = {} & \frac{a_2}{a_1 - \alpha} - \frac{a_3}{(a_1 - \alpha)^2} + \\ & + \frac{a_4}{(a_1 - \alpha)^3} - \frac{a_5}{(a_1 - \alpha)^4} \end{aligned}$	$\begin{aligned} & (x + a_1 - \alpha) \times \\ & \times \Big\{ x^4 + \alpha x^3 + [a_2 - \alpha(a_2 - \alpha)]\, x^2 + \\ & + \{a_3 - (a_1 - \alpha)\, [a_2 - \alpha(a_1 - \alpha)]\}\, x + \\ & + \frac{a_5}{a_1 - \alpha} \Big\} \end{aligned}$
$\begin{aligned} \alpha = {} & \frac{a_3}{a_2 - \beta} - \frac{a_5 + a_4(a_1 - \alpha)}{(a_2 - \beta)^2} + \\ & + \frac{(a_1 - \alpha)^2\, a_5}{(a_2 - \beta)^3}\ ; \\ \beta = {} & \alpha(a_1 - \alpha) + \\ & + \frac{a_4}{a_2 - \beta} - \frac{(a_1 - \alpha)\, a_5}{(a_2 - \beta)^2} \end{aligned}$	$\begin{aligned} & [x^2 + (a_1 - \alpha)\, x + a_2 - \beta] \times \\ & \times \Big\{ x^3 + \alpha x^2 + \\ & + [\beta - \alpha(a_1 - \alpha)] \times \\ & \times x + \frac{a_5}{a_2 - \alpha} \Big\} \end{aligned}$
$\begin{aligned} \alpha &= \frac{a_6}{a_5 - \eta}\ ; \\ \beta &= \alpha(a_1 - \alpha); \\ \gamma &= \alpha(a_2 - \beta); \\ \delta &= \alpha(a_3 - \gamma); \\ \eta &= \alpha(a_4 - \delta) \end{aligned}$	$\begin{aligned} & (x + \alpha)\, [x^5 + (a_1 - \alpha)\, x^4 + \\ & + (a_2 - \beta)\, x^3 + (a_3 - \gamma)\, x^2 + \\ & + (a_4 - \delta)\, x + a_5 - \eta] \end{aligned}$
$\begin{aligned} \alpha &= \frac{a_5}{a_4 - \delta} - \frac{a_6(a_3 - \gamma)}{(a_4 - \delta)^2}\ ; \\ \beta &= \alpha(a_1 - \alpha) + \frac{a_6}{a_4 - \delta}\ ; \\ \gamma &= \alpha(a_2 - \beta) + \frac{a_6}{a_4 - \delta}(a_1 - \alpha); \\ \gamma &= \alpha(a_3 - \gamma) + \frac{a_6}{a_4 - \delta}(a_2 - \beta) \end{aligned}$ or other formulae, obtained from the system of equations (4.6)	$\begin{aligned} & \left(x^2 + \alpha x + \frac{a_6}{a_4 - \delta}\right) \times \\ & \times [x^4 + (a_1 - \alpha)\, x^3 + \\ & + (a_2 - \beta)\, x^2 + \\ & + (a_3 - \gamma)\, x + a_4 - \delta] \end{aligned}$

Often the transient process is determined with great accuracy from the first two or three roots, and then the iterative method of calculation of the roots proves to be very economical.

(c) *General remarks about the method*

When all the parameters of the system are chosen and the determination of the roots of the equation is only a verification stage in the calculation, the method described above, after some practice, may be performed quickly. But when a constructed process proves to be unsatisfactory, the method described does not give any advice as to how the parameters of the system must be changed in order to improve the control process. Indeed, the equation of the process, (4.4), depends in a complicated way upon the roots of the characteristic equation, the dependence of these roots on the coefficients of the equation being unknown, and these coefficients themselves depend in a complicated way on the parameters of the given system. This makes it necessary to seek ways of constructing the control process which do not involve calculating the roots of the characteristic equation.

3. A Graphical Method of Constructing the Control Process

The control process can be constructed without a preliminary calculation of the roots of the characteristic equation with the help of various graphical methods of approximate integration of the control process equations.

The singularity of control systems consists in the peculiar "chain" structure of the system of equations describing the process : each element of the system acts on the following element or on an inner loop which itself consists of a sequential chain of elements. In order to construct the processes in systems having a similar structure, it is convenient to use a graphical method based on a known method for constructing exponential functions. A description of this method is first made in the construction of an exponential curve, and then the construction of processes in separate elements of the control chain is described, and, finally, in the chain as a whole.

(a) *The construction of exponential functions*

The method we describe for constructing the exponential curve is based on the following properties of this function.

Property I. The ratios of the ordinates of points equidistant from one another are equal.

Let us assume that we are given the exponential function

$$x = C' e^{-\frac{t}{T}},$$

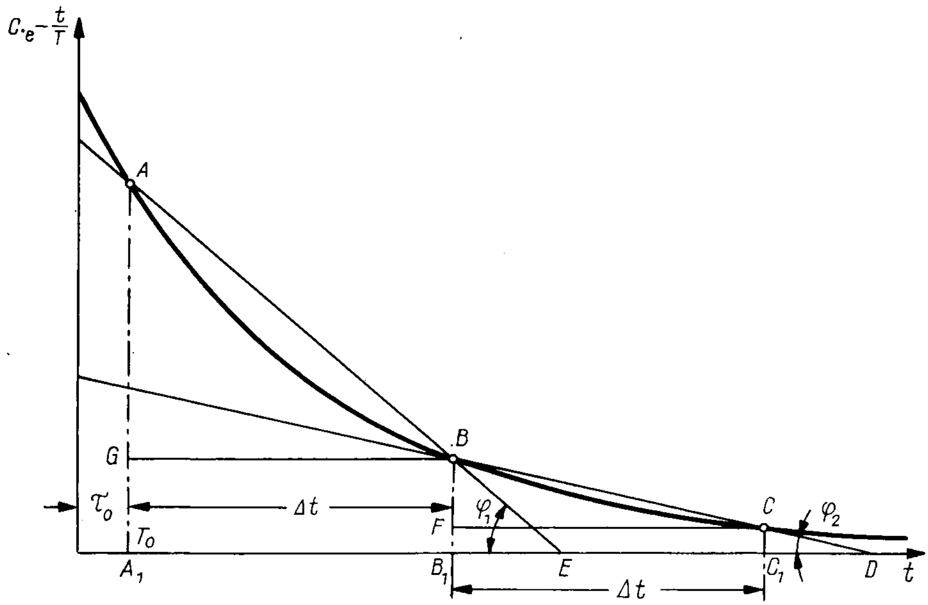


Fig. 146

Let $A_1 B_1 = B_1 C_1$ (Fig. 146), i.e. if the abscissa of the point A is equal to τ_0, then the abscissa of the point B is equal to $\tau_0 + \Delta t$, and of the point C to $\tau_0 + 2 \Delta t$. Let us draw a secant through A and B, and also through B and C. We find the ordinates of the points A, B and C:

$$AA_1 = C' e^{-\frac{\tau_0}{T}},$$

$$BB_1 = C' e^{-\frac{\tau_0 + \Delta t}{T}},$$

$$CC_1 = C' e^{-\frac{\tau_0 + 2\Delta t}{T}}.$$

Calculating the ratios $\frac{AA_1}{BB_1}$ and $\frac{BB_1}{CC_1}$ we obtain

$$\frac{AA_1}{BB_1} = \frac{BB_1}{CC_1} = e^{\frac{\Delta t}{T}} = \text{const.}$$

Property II. Secants drawn through points equidistant from one another have equal projections.

On the basis of Property I, we have the equation

$$\frac{AA_1}{BB_1} = \frac{BB_1}{CC_1}.$$

But

$$BB_1 = AA_1 - AG$$

and

$$CC_1 = BB_1 - BF.$$

From the triangles ABG and BCF we have

$$AG = \Delta t \tan \varphi_1$$

and

$$BF = \Delta t \tan \varphi_2 ,$$

where φ_1 is the angle between the first secant and the x-axis, and φ_2 is the angle between the second secant and the x-axis. But

$$\tan \varphi_1 = \frac{AA_1}{A_1 E}$$

and

$$\tan \varphi_2 = \frac{BB_1}{B_1 D}.$$

Hence

$$\frac{AA_1}{BB_1} = \frac{BB_1}{CC_1} = \frac{AA_1}{AA_1 - AG} = \frac{BB_1}{BB_1 - BF} =$$

$$= \frac{AA_1}{AA_1 - \frac{AA_1}{A_1 E}\Delta t} = \frac{BB_1}{BB_1 - \frac{BB_1}{B_1 D}\Delta t},$$

whence we obtain

$$1 - \frac{\Delta t}{A_1 E} = 1 - \frac{\Delta t}{B_1 D},$$

or

$$A_1 E = B_1 D .$$

Property III. The projection drawn through two neighbouring points is approximately equal to

$$T + \frac{\Delta t}{2} .$$

We determine the length of the projection.

From the similar triangles ABG and $AA_1 E$ we can write the following equation:

$$\frac{AA_1}{A_1 E} = \frac{AG}{\Delta t},$$

which gives

$$A_1 E = \frac{AA_1}{AG} \Delta t = \frac{C' e^{-\frac{\tau_0}{T}} \Delta t}{C' e^{-\frac{\tau_0}{T}} - C' e^{-\frac{\tau_0 + \Delta t}{T}}} = \frac{\Delta t}{1 - e^{-\frac{\Delta t}{T}}} .$$

The right-hand side of this equation can be expanded in a series of ascending powers of Δt :

$$A_1 E = T + \frac{\Delta t}{2} \left[1 + \frac{1}{6} E - \frac{2}{6!} E^3 - \frac{2}{7!} E^5 - \dots \right],$$

where

$$E = \frac{\Delta t}{T} .$$

Restricting ourselves to the first two terms of the expansion, we obtain a final formula which determines approximately the length of the projection :

$$A_1 E = \frac{\Delta t}{1 - e^{-\frac{\Delta t}{T}}} \approx T + \frac{\Delta t}{2} .$$

The smaller the selected intervals Δt, the more accurate the result will be.

In practice it is sufficient to take

$$\Delta t = \left(\frac{1}{5} \text{ to } \frac{1}{10}\right) T.$$

We can now proceed to describe the construction of the exponential function.

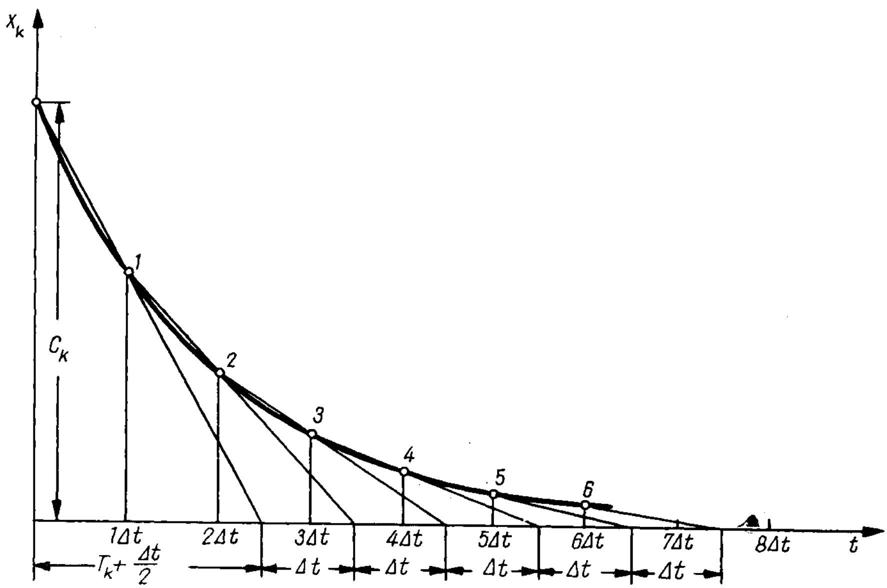


FIG. 147

Let us assume that it is necessary to construct the exponential

$$x_k = C_k e^{-\frac{t}{T}}.$$

We will make the construction in Cartesian coordinates. Along the ordinate axis we put x_k, and as abscissa we take t.

Let us divide the t-axis into intervals equal to Δt, such that $\Delta t \leqslant T_k$ and $T_k = n \Delta t$, i.e. so that T_k is a multiple of the interval Δt. We halve each of the resulting intervals.

We begin the construction by measuring a segment equal to C_k along the x_k-axis from the origin, and along the t-axis, the segment (Fig. 147).

$$T_k + \frac{\Delta t}{2}.$$

The two points we obtain are joined by a straight line. From the point with abscissa $1\,\Delta t$ we produce a line parallel to the y-axis, to intersect the straight line we have just constructed.

The point of intersection, *1*, which we obtain will be a point of the required exponential.

Then, along the t-axis we mark off the segment

$$T_k + 3\frac{\Delta t}{2}.$$

Through this point and the point of the exponential found in the previous construction we draw a second straight line. From the point with abscissa $2\,\Delta t$ we produce a straight line parallel to the x_k-axis, and so find a second point of the exponential. Repeating the construction, we find all the remaining points.

An example of this construction is shown in Fig. 147.

(b) *The construction of the transient process for one single-capacitance stage*

We consider a single-capacitance stage, at whose input a unit disturbance $A \cdot \mathbf{1}$ acts (Fig. 148).

The differential equation is written in the form :

$$T\,\frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = k \cdot A \cdot I.$$

Integrating this equation, we obtain the equation of the transient process for the output coordinate of a single-capacitance stage

$$x_{\text{out}} = kA\left(1 - e^{-\frac{t}{T}}\right).$$

For $t = 0$ $x_{\text{out}} = 0$, and for $t = \infty$ $x_{\text{out}} = kA = A_1$.

Removing the brackets on the right-hand side, and taking the quantity A_1 to the left-hand side, we can rewrite the equation of the transient process in the following form :

$$x_{\text{out}} - A_1 = - A_1 e^{-\frac{t}{T}}.$$

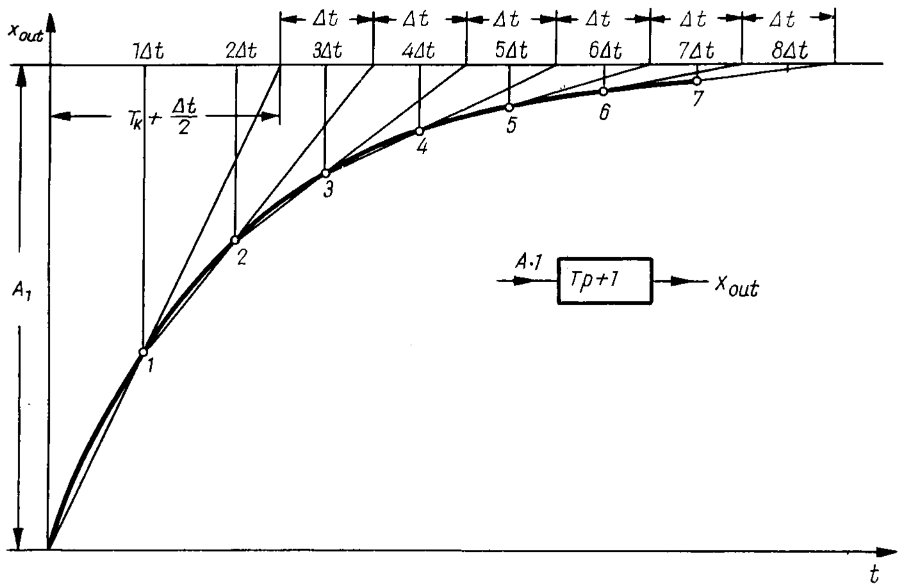


FIG. 148

Hence, in order to construct the transient process we must construct the exponential function

$$\tilde{x}_{\text{out}} = - A_1 e^{-\frac{t}{T}}$$

and then shift the ordinate axis to the point A_1.

It is convenient to shift the axis from the very start. To do this we cut off a segment equal to A_1 on the x_{out}-axis (Fig. 148). Through its end-point we draw a line parallel to the t-axis and from A_1 mark off a segment equal to $T + \frac{\Delta t}{2}$ · on it. We join the resulting point to the origin of coordinates. From the point with abscissa $1\,\Delta t$ we

produce a line parallel to the x_{out}-axis to meet the earlier constructed line. Their point of intersection, *1*, will be a point of the required exponential. On the straight line through A_1 we mark off the segment $T + 3\,\frac{\Delta t}{2}$ from A_1. We join the end-point of the exponential which we have just found by a straight line to the previous point on the exponential, and then from the point with abscissa $2\,\Delta t$ we produce a straight line parallel to the x_{out}-axis. The point of intersection of these two lines will be another point of the exponential. Continuing in this way we find its other points. This construction is shown in Fig. 148.

(c) *The construction of the process for the output coordinate of a single-capacitance stage with an arbitrary disturbance $f(t)$ at its input*

We assume that an arbitrary disturbance $f(t)$ acts at the input of a single-capacitance stage. Let this function be given by a graph.

We divide the t-axis into segments Δt such that $\Delta t \ll T$ and $T = n\,\Delta t$. In addition, these intervals Δt must be small enough for

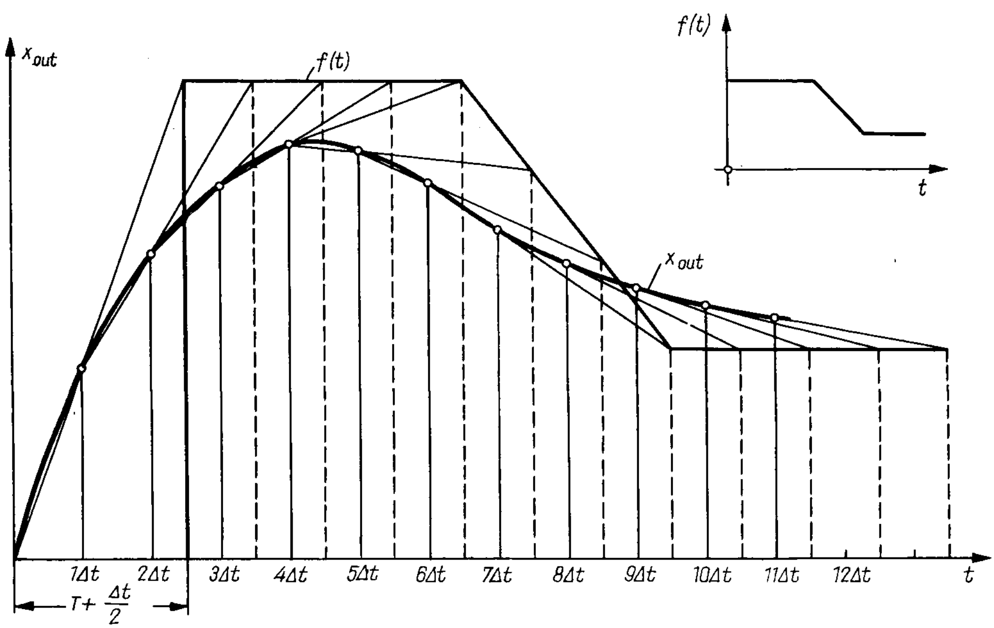


FIG. 149

$f(t)$ to be represented with a sufficient degree of accuracy by a broken line (Fig. 149). We shift the graph of $f(t)$ along the t-axis by an amount $T + \frac{\Delta t}{2}$.

We join the first point of the curve by a straight line to the origin of coordinates and by the construction given above find on it a point belonging to the required transient process. We find the next point of $f(t)$ at a distance Δt from the first, and joining it to the point found previously by the same construction, we find the next point of the curve of the transient process. Repeating this construction, we find the whole curve.

An example of this construction is given in Fig. 149.

(d) *The construction of the transient process for an open network of single-capacitance stages with a unit disturbance $A \cdot 1$ at its input*

Let us consider an open circuit of single-capacitance stages, at whose input the unit disturbance $A \cdot 1$ is applied (Fig. 150*a*).

The construction of the transient process for the first stage differs in no way from the earlier construction for a single-capacitance stage.

The construction is as follows :

1. We construct the transient process for the first stage of the circuit in the same way as if it were by itself.

2. We use the resulting curve of the change in the output coordinate of the first stage as the disturbance for the second stage.

3. In the same way we repeat the construction for the second stage. The deviation of the output coordinate of the first stage at the end of the lth interval becomes the disturbance for the second stage for the whole of the $(l + 1)$th interval.

A similar construction is made for all the remaining stages.

This construction can be made at the same time for all the stages for any one interval.

Let the deviations of all the coordinates in the lth interval be known. Then, to find the deviations in the $(l + 1)$th interval, we make the construction described above, taking the value of the external disturbance at the end of the lth interval as the action on the first stage.

The output coordinate of the first stage x_1 at the end of the lth interval acts on the second stage.

The output coordinate of the second stage at the end of the lth interval acts on the third stage, and so on.

The construction is given in Fig. 150*b* for a circuit consisting of three single-capacitance stages.

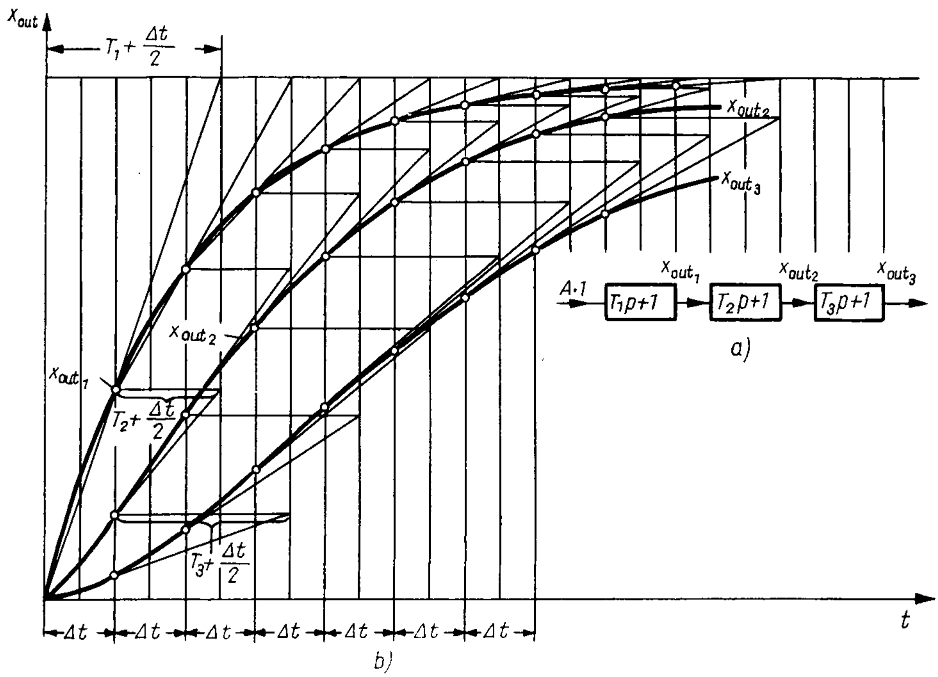


FIG. 150

(e) *The construction of the transient process for a closed circuit of single-capacitance stages when the unit disturbance* $f(t) = 1$ *is applied at its input*

Let us now consider a closed circuit of single-capacitance stages, at whose input the unit disturbance $f(t) = 1$ is applied.

For all the intermediate stages, the construction of the transient process will be made according to the rules developed above.

The only difference is in the construction of the transient process for the output coordinate of the first stage.

As the disturbance of the first stage for the $(l+1)$th interval, we must take the difference between the magnitude of the external disturbance at the end of the lth interval and the magnitude of the output coordinate of the circuit at the end of the $(l-1)$th interval.

An example of this construction is given in Fig. 151 for the case when $f(t) = A \cdot \mathbf{1}$.

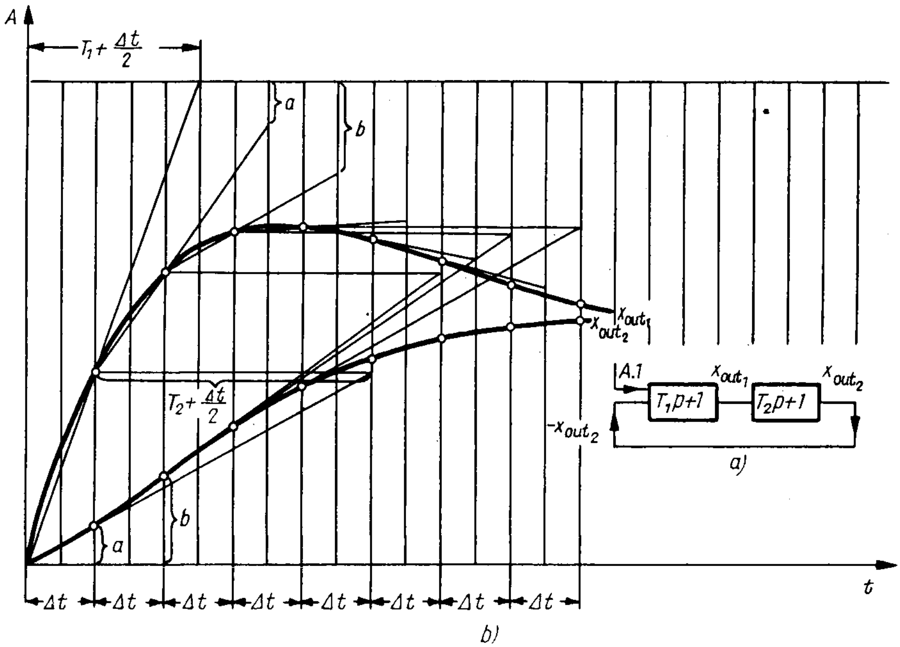


FIG. 151

(f) *The construction of the transient process for an astatic stage*

We now pass to the consideration of the astatic stage

$$T\frac{dx_{out}}{dt} = x_{in}.$$

To both sides of this equation we add x_{out}:

$$T\frac{dx_{out}}{dt} + x_{out} = x_{in} + x_{out}.$$

Then the astatic stage can be considered as a single-capacitance stage, embraced by positive feedback (Fig. 152*a*).

The construction of the transient process from the output coordinate of an astatic stage is carried out just as if the stage were single-capacitance, but we regard not only the output coordinate of the previous stage, but also the output coordinate of the given

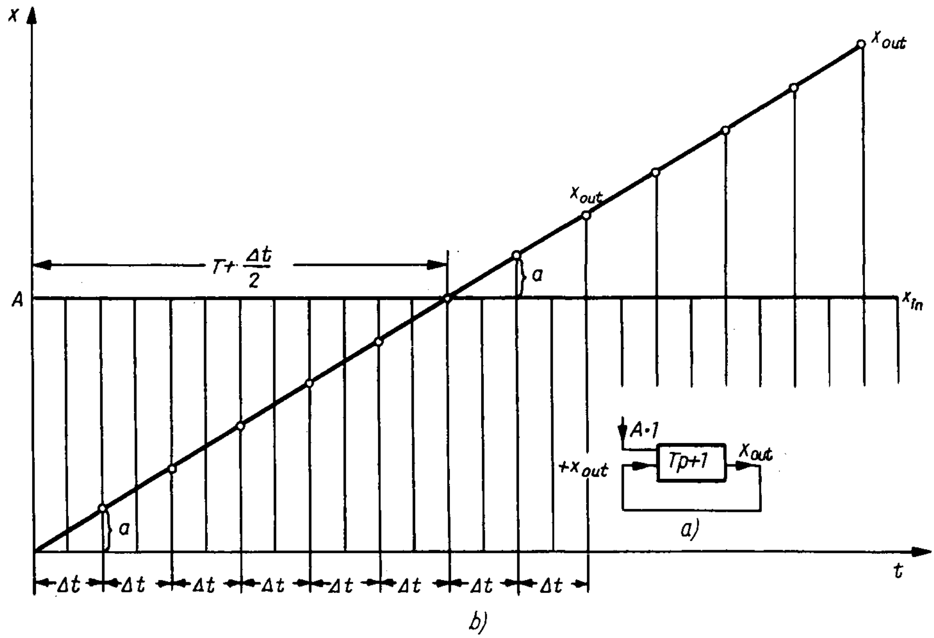


FIG. 152

astatic stage at the end of the previous interval, as acting at the input of this stage.

Thus, to construct the value of the output coordinate of an astatic stage for the $(l+1)$th interval, it is necessary to sum the value of the output coordinate of the preceding stage at the end of the lth interval and the value of the output of the latter stage at the end of the same interval. The point found by this construction is joined by a straight line to the point determined by the construction for the lth interval and by the usual construction we obtain the value of the output coordinate for the $(l+1)$th interval.

An example of this construction is given in Fig. 152*b*.

(g) *The construction of the process in an oscillatory stage*

We assume that the given circuit contains an oscillatory stage (Fig.153a) :

$$T'^2 \frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + T_k \frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + x_{\text{out}} = kx_{\text{in}}.$$

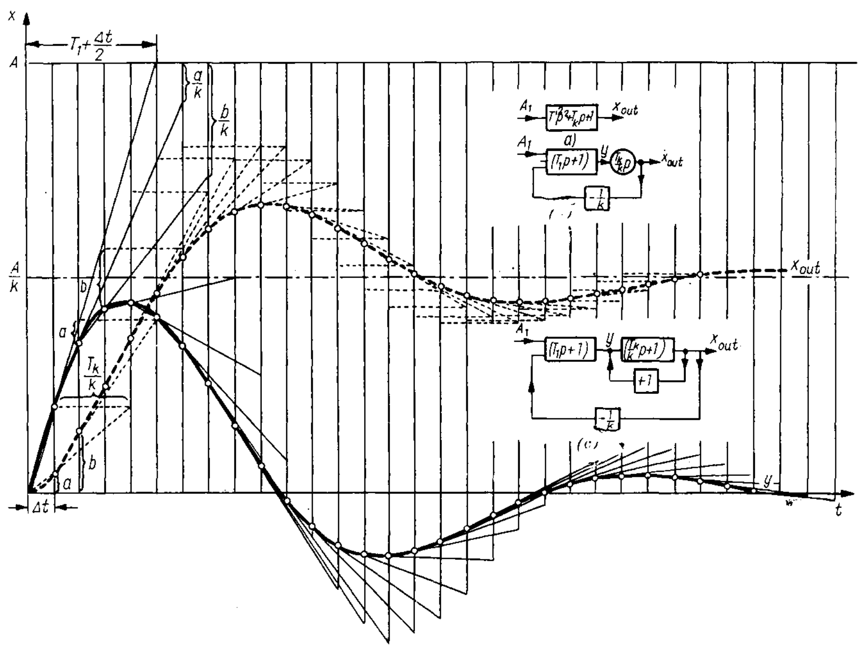


FIG. 153

We divide both sides of the equation by k and put

$$T'^2 = T_1 T_k.$$

Then

$$\frac{T_1 T_k}{k} \frac{\mathrm{d}^2 x_{\text{out}}}{\mathrm{d}t^2} + \frac{T_k}{k} \frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} + \frac{1}{k} x_{\text{out}} = x_{\text{in}},$$

or

$$T_1 \frac{\mathrm{d}y}{\mathrm{d}t} + y = x_{\text{in}} - \frac{1}{k} x_{\text{out}},$$

where

$$\frac{T_k}{k} \frac{\mathrm{d}x_{\text{out}}}{\mathrm{d}t} = y.$$

Hence, the oscillatory stage can be replaced by a circuit consisting of two stages: one astatic and one single-capacitance, embraced by negative feedback (Fig. 153*b*), or by two single-capacitance stages with time constants $T_1 = \frac{T'^2}{T_k}$ and $\frac{T_k}{k}$ embraced by negative feedback with a coefficient of amplification $\frac{1}{k}$, the stage with time constant $\frac{T_k}{k}$ in its turn being closed with positive feedback with a coefficient of amplification 1 (Fig. 153*c*). In the construction we take into account that (because of feebdack) the difference between the output coordinate of the preceding stage and the output coordinate of the second single-capacitance stage, decreased k times, acts at the intput of the first single-capacitance stage, and that the sum of the output coordinates of the first stage for the lth interval and the output coordinate of the second stage for the $(l - 1)$th interval acts at the input of the second single-capacitance stage.

The construction of the transient process in the case when a disturbance $A \cdot 1$ is applied at the input of the first stage is shown in Fig. 153.

(h) *Additional remarks*

The construction we have described is widely used in the same way in systems containing unstable elements, more complicated internal loops and derivative action. Any loop and any elements can be represented by a corresponding number of single-capacitance stages and by supplementary internal positive and negative feedback, and the construction of the process in a system containing loops, consisting of single-capacitance stages, for any disturbance given by a graph, is clear from the above description.

In conclusion we note that the given graphical method for constructing transient processes enables us also to take some nonlinearities existing in the given system of automatic control into account*.

* A method similar to this is explained in: Popov, Ye. P., The Dynamics of Automatic Control Systems (Dinamika sistem Avtomaticheskogo regulirovaniya), Gostechizdat (1954).

4. The Construction of the Control Process from the Frequency Characteristics of the System

(a) *General remarks*

In cases when the properties of the system are given by its frequency characteristics, the control process for any given disturbance can be constructed, starting from the frequency characteristics, without calculating the roots of the characteristic equation. For this purpose we find, first of all, the Fourier transform of the given coordinate of the system for the given disturbance.**

Suppose that it is required to construct the change as a function of time of any coordinate (we shall call it the coordinate x, omitting any suffix) under the action of a disturbance $f(t)$, applied at some point of the system. We denote by $\Phi_x^*(i\omega)$ the Fourier transform of the required function $x(t)$. Then

$$\Phi_x^*(i\omega) = \Phi(i\omega)\,\Phi_{f(t)}(i\omega)\,, \tag{4.9}$$

where $\Phi(i\omega)$ is the frequency characteristic of the considered closed loop system from the point of application of the action $f(t)$ to the coordinate x, $\Phi_{f(t)}(i\omega)$ is the Fourier transform or spectrum of the action $f(t)$:

$$\Phi_{f(t)}(i\omega) = \int_0^\infty f(t)\,e^{-i\omega t}\,\mathrm{d}t\,. \tag{4.10}$$

We recall that we obtain $\Phi(i\omega)$ and $\Phi_{f(t)}(i\omega)$ by replacing p by $i\omega$ in the transfer function $W(p)$ of the system, and in the Laplace transform of the function $f(t)$, respectively. But in contrast to the integral defining the Laplace transform, the integral $\int_0^\infty f(t)\,e^{-i\omega t}\,\mathrm{d}t$ has a meaning only if the function $f(t)$ satisfies the condition $\lim_{t\to\infty} f(t) = 0$. We assume, moreover, that the function $f(t)$ is bounded and continuous for all $t \geqslant 0$. Actions which satisfy this condition are called *vanishing* as distinct from *non-vanishing* actions which, remaining continuous and bounded for all $t \geqslant 0$, tend to a non-zero limit as $t \to \infty$. As we shall show later, the Fourier transform can also be used for some non-vanishing actions, but this requires special consideration.

** See Appendix I (p. 500).

If the Fourier transform $\Phi_x^*(i\omega)$ for the coordinate x is calculated, the required function $x(t)$ is then determined by the inverse Fourier transform

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_x^*(i\omega)\, e^{i\omega t}\, d\omega. \tag{4.11}$$

(b) *The construction of the process in the case of a vanishing disturbance*

If $\lim_{t\to\infty} f(t) = 0$, i.e. if the disturbance is vanishing, then the process can be constructed immediately from equation (4.11). For this purpose it is convenient to transform it as follows.

We separate real and imaginary parts in $\Phi_x^*(i\omega)$

$$\Phi_x^*(i\omega) = P^*(\omega) + i\,Q^*(\omega)\,,$$

and use Euler's identity

$$e^{i\omega t} = \cos\omega t + i \sin\omega t\,.$$

Putting these values in the product $\Phi_x^*(i\omega)\, e^{i\omega t}$ we obtain

$$\Phi_x^*(i\omega)\, e^{i\omega t} = [P^*(\omega)\cos\omega t - Q^*(\omega)\sin\omega t] + \\ + i[Q^*(\omega)\cos\omega t + P^*(\omega)\sin\omega t]\,.$$

Then

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} [P^*(\omega)\cos\omega t - Q^*(\omega)\sin\omega t]\, d\omega + \\ + \frac{i}{2\pi}\int_{-\infty}^{+\infty} [Q^*(\omega)\cos\omega t + P^*(\omega)\sin\omega t]\, d\omega.$$

The function $P^*(\omega)$ is even, and the function $Q^*(\omega)$ is odd. The integrand of the second integral is therefore odd and this integral is equal to zero.

$$\int_{-\infty}^{\infty} [Q^*(\omega)\cos\omega t + P^*(\omega)\sin\omega t] \equiv 0. \tag{4.12}$$

The real part of the above equation, therefore, defines $x(t)$:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [P^*(\omega) \cos \omega t - Q^*(\omega) \sin \omega t] \, d\omega. \tag{4.13}$$

Since we are only considering those functions $f(t)$ which are equal to zero for all $t < 0$ then $x(t) = 0$ for $t < 0$. Taking t to be positive, we replace t by $-t$ in (4.13) and equate $x(t)$ to zero.

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} [P^*(\omega) \cos(\omega t) + Q^*(\omega) \sin(\omega t)] \, d\omega = 0. \tag{4.14}$$

Adding equations (4.14) and (4.13) we find

$$x(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} P^*(\omega) \cos \omega t \, d\omega. \tag{4.15}$$

But $P^*(\omega)$ is an even function. Therefore

$$\int_{-\infty}^{+\infty} P^*(\omega) \cos \omega t \, d\omega = 2 \int_{0}^{+\infty} P^*(\omega) \cos \omega t \, d\omega,$$

and the integral (4.15) can be written in the following form :

$$\boxed{x(t) = \frac{2}{\pi} \int_{0}^{+\infty} P^*(\omega) \cos \omega t \, d\omega} \tag{4.16}$$

Subtracting (4.14) from (4.13) we obtain similarly

$$\boxed{x(t) = -\frac{2}{\pi} \int_{0}^{\infty} Q^*(\omega) \sin \omega t \, d\omega.} \tag{4.17}$$

The equation (4.16) defines the function $x(t)$ in terms of the real part $P^*(\omega)$ of the Fourier transform of this function, which is equal to the product of the frequency characteristic of the system, $W(i\omega)$, and the Fourier transform of the acting disturbance (i.e. its complex spectrum).

We recall that the above reasoning is true only if the given system is stable, if $f(t)$ and $x(t)$ are zero for $t < 0$ and if the disturbance is vanishing.

The problem is now reduced to the calculation of the integral (4.16) with the help of some approximate methods. We restrict ourselves to the description of two methods suitable for this purpose.

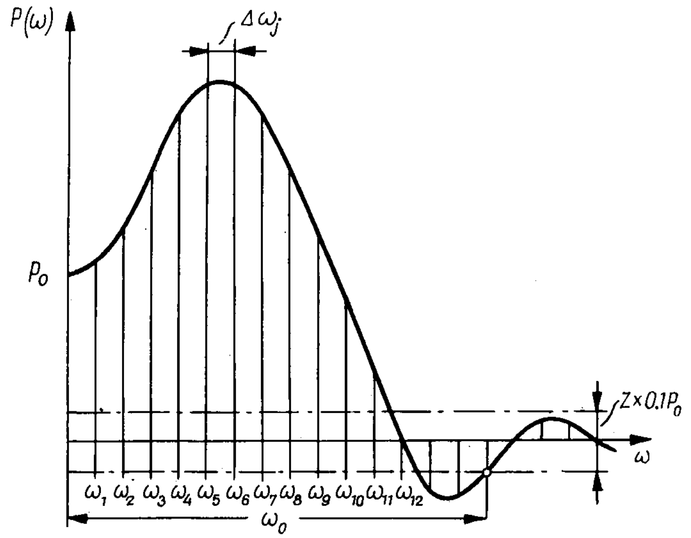


FIG. 154

1st Method. We divide the ω-axis into intervals $\Delta\,\omega$ (they can be unequal) and for each point of division we find the ordinate $P^*(\omega)$. We note that $P^*(\omega_j)$ usually tends to zero as $\omega \to \infty$. It is not usually necessary to take the whole range of the change in ω. We take the point ω_0 after which all $P_j^* < 0{\cdot}1\, P_0^*$, as the boundary point (Fig. 154). The integration will be made from zero to ω_0 for all the intervals $\Delta\,\omega_j$:

$$x(t) = \frac{2}{\pi}\int\limits_0^{\omega_1} P_1^* \cos\omega t\,\mathrm{d}\omega + \frac{2}{\pi}\int\limits_{\omega_1}^{\omega_2} P_2^* \cos\omega t\,\mathrm{d}\omega + \dots$$

$$\dots + \frac{2}{\pi}\int\limits_{\omega_{n-1}}^{\omega_n} P_n^* \cos\omega t\,\mathrm{d}\omega .$$

Therefore

$$x(t) = \frac{2}{\pi t}\left[P_3^* (\sin \omega_1 t - \sin 0 \cdot t) + P_2^* (\sin \omega_2 t - \sin \omega_1 t) + \right.$$
$$+ P_3^* (\sin \omega_3 t - \sin \omega_2 t) + P_4^* (\sin \omega_4 t - \sin \omega_3 t) + \dots$$
$$\left. + P_n^* (\sin \omega_n t - \sin \omega_{n-1} t)\right].$$

Collecting similar terms, we obtain :

$$x(t) = \frac{2}{\pi t}\left[(P_1^* - P_2^*) \sin \omega_1 t + (P_2^* - P_3^*) \sin \omega_2 t + \right.$$
$$+ (P_3^* - P_4^*) \sin \omega_3 t + (P_4^* - P_5^*) \sin \omega_4 t + \dots$$
$$\left. + (P_{n-1}^* - P_n^*) \sin \omega_{n-1} t + P_n^* \sin \omega_n t\right].$$

In using this method to construct the process we need only use tables of trigonometric functions, but on the other hand, to construct one point we have to calculate a large number of terms (sometimes from 30 to 40) in order to represent the function $P^*(\omega)$, by a series of discrete steps, with the required accuracy. We can considerably reduce the number of terms if we replace the curve $P^*(\omega)$ not by a step-function, but by a broken straight line with sloping sections.

2nd Method. We replace the curve $P^*(\omega)$ by a broken straight line, describing the curve sufficiently well (for example, the line *chbfe* in Fig. 155*a*). We project the joins *c*, *h*, *b*, *f*, *e* etc. on to the y-axis. The links of the line and the projecting straight lines parallel to the x-axis then form trapezia and triangles, one of whose sides lies on the y-axis. The area between the broken straight lines and the x-axis can be obtained by summing the areas of these trapezia and triangles, taking the sign into account. Thus, the area bounded by the broken line *chbfe* for example, shown in Fig. 155*a*, can be obtained by subtracting the area of the trapezium *dgfe* and that of the triangle *ahc* from the area of the trapezium *abfg*. Of course, these areas are not altered if we displace the trapezia and triangles so that one of their vertices lies at the origin of coordinates, one of their sides lies along the x-axis and another side along the y-axis (Fig. 155*b*).

As a result, the integral of $P^*(\omega)$ can be approximated by the integral of the broken straight line constructed to describe the curve $P^*(\omega)$ sufficiently well, and this integral in its turn can be replaced

by the sum or difference of the integrals of the curves bounding trapezia or triangles with one vertex at the origin of coordinates and with sides such that one lies along the x-axis and another along the y-axis.

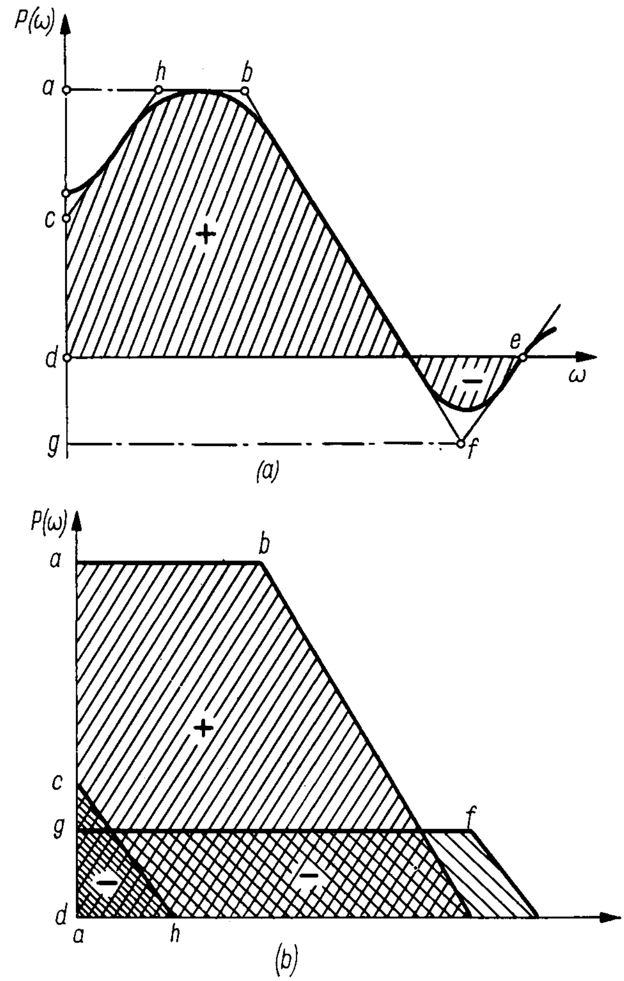


FIG. 155

Thus, in the case of Fig. 155*b*

$$\int_0^\infty P^*(\omega)\cos\omega t\,\mathrm{d}\omega \approx \int (chbfe)\cos\omega t\,\mathrm{d}\omega = \int (\diagup\!\overline{\ \ }\!\diagdown\, gabf)\cos\omega t\,\mathrm{d}\omega -$$
$$- \int (\diagup\!\overline{\ \ }\!\diagdown\, dgfe)\cos\omega t\,\mathrm{d}\omega - \int (\Delta ach)\cos\omega t\,\mathrm{d}\omega\,.$$

Here, in the brackets under the integral signs we use the conventional notation for the broken straight line (Fig. 155*a*) whose equation and corresponding limits of integration must be substituted in the integral.

The triangles with one vertex at the origin of coordinates and with two sides lying on the co-ordinate axes can be considered as a particular case of a trapezium whose base is equal to zero.

The problem reduces to the calculation of the integrals for the trapezia. We denote by $\lambda(\omega)$ the equation of the curve bounding one of these trapezia.

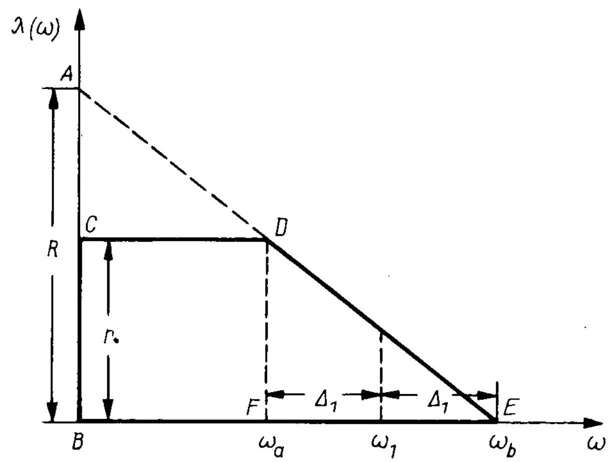


Fig. 156

Then, on the strength of this reasoning we have

$$x(t) \approx \frac{2}{\pi} \sum_{j=1}^{j=n} \int_0^\infty \lambda_j(\omega) \cos \omega t \, d\omega , \qquad (4.18)$$

where the summation is made over all the trapezia. Let us calculate the integral

$$\int_0^\infty \lambda_j(\omega) \cos \omega t \, d\omega$$

for one of the trapezia.

Consider the trapezium $BCDE$ (Fig. 156).

The required integration will be performed along two straight lines: along CD for ω going from zero to ω_a, and along DE for ω from ω_a to ω_b.

The equation of the first straight line is $\lambda(\omega) = r = \text{const.}$ We find the equation of the second line DE.

From the triangles ABE and DEF we have :

$$\frac{R}{\omega_1 + \Delta_1} = \frac{r}{2\Delta_1},$$

whence

$$R = \frac{r(\omega_1 + \Delta_1)}{2\Delta_1},$$

where

$$\Delta_1 = \frac{\omega_b - \omega_a}{2}, \quad \omega_1 = \frac{\omega_b + \omega_a}{2}.$$

Thus, the equation of DE is

$$\lambda(\omega) = -\frac{r}{2\Delta_1}\omega + \frac{r(\omega_1 + \Delta_1)}{2\Delta_1},$$

or

$$\lambda(\omega) = \frac{r}{2\Delta_1}(\omega_1 + \Delta_1 - \omega),$$

for values of ω such that

$$\omega_1 - \Delta_1 \leqslant \omega \leqslant \omega_1 + \Delta_1.$$

We calculate the integral :

$$\int_0^\infty \lambda(\omega)\cos\omega t\,\mathrm{d}\omega = r\int_0^{\omega_1-\Delta_1}\cos\omega t\,\mathrm{d}\omega +$$

$$+\frac{r}{2\Delta_1}\int_{\omega_1-\Delta_1}^{\omega_1+\Delta_1}(\omega_1+\Delta_1-\omega)\cos\omega t\,\mathrm{d}\omega = r\int_0^{\omega_1-\Delta_1}\cos\omega t\,\mathrm{d}\omega +$$

$$+\frac{r}{2\Delta_1}(\omega_1+\Delta_1)\int_{\omega_1-\Delta_1}^{\omega_1+\Delta_1}\cos\omega t\,\mathrm{d}\omega - \frac{r}{2\Delta_1}\int_{\omega_1-\Delta_1}^{\omega_1+\Delta_1}\omega\cos\omega t\,\mathrm{d}\omega =$$

$$= \frac{r}{t}\sin\omega t\Big|_0^{\omega_1-\Delta_1} + \frac{r}{2\Delta_1 t}(\omega_1+\Delta_1)\sin\omega t\Big|_{\omega_1-\Delta_1}^{\omega_1+\Delta_1} -$$

$$-\frac{r}{2\Delta_1}\left(\frac{1}{t^2}\cos\omega t+\frac{\omega}{t}\sin\omega t\right)\Bigg|_{\omega_1-\Delta_1}^{\omega_1+\Delta_1}=$$

$$=-\frac{r}{2\Delta_1 t^2}\left[\cos(\omega_1+\Delta_1)t-\cos(\omega_1-\Delta_1)t\right]=$$

$$=\frac{r\omega_1}{\Delta_1 t^2\omega_1}\sin\omega_1 t\sin\Delta_1 t.$$

But $r\,\omega_1 = A$, the area of the trapezium.

Therefore

$$\int_0^\infty \lambda(\omega)\cos\omega t\,d\omega = A\,\frac{\sin\omega_1 t}{\omega_1 t}\,\frac{\sin\Delta_1 t}{\Delta_1 t}. \tag{4.19}$$

We recall (see Fig. 156) that in (4.19)

$$\omega_1=\frac{\omega_a+\omega_b}{2}, \quad\text{and}\quad \Delta_1=\frac{\omega_b-w_a}{2}.$$

Putting (4.19) in (4.18) we obtain finally :

$$\boxed{x_j(t)=\frac{2}{\pi}\sum A\left(\frac{\sin\omega_1 t}{\omega_1 t}\right)\left(\frac{\sin\Delta_t t}{\Delta_1 t}\right)}. \tag{4.20}$$

The summation in (4.20) is made over all the trapezia.

For convenience in calculating processes with the use of this formula, a table of the values of the function $\frac{\sin x}{x}$ is given at the end of this book*.

Let us now summarize what has been said in this section, listing the operations which we have to perform in determining the co-ordinate x as a function of time for a system under the action of vanishing action $f(t)$ (i.e. which has the property $\lim_{t\to\infty} f(t) = 0$. In order to construct the process it is necessary :

1. To compute the frequency characteristic of the closed system from the point of application of the disturbance $f(t)$ to the coordinate $x_j = x$.

* See Appendix 2 (p. 505).

2. To compute the complex spectrum (i. e. the Fourier transform) of the acting disturbance $f(t)$.

3. To form the product of the complex functions found in 1 and 2 and to find its real part $P^*(\omega)$.

4. To draw the graph of the scalar function $P^*(\omega)$ and to replace it by a broken straight line.

5. To project the joins of this broken line on to the γ-axis and to construct trapezia (or, in special cases, triangles) with one vertex at the origin of coordinates, and two sides lying along the co-ordinate axes, the sum or difference of their areas being equal to the area bounded by the broken line, taking sign into account.

6. To calculate from formula (4.20) with the aid of the tables given in Appendix 2 points of the process for various values of t.

(c) *The construction of the process for a non-vanishing disturbance*

We consider now the case when a non-vanishing disturbance $\mathbf{1}f(t)$ acts on the system, i.e. one such that $\lim\limits_{t\to\infty} \mathbf{1}f(t) = A \neq 0$.

We form the function $\varphi(t) = \mathbf{1}f(t) - A \cdot \mathbf{1}$. This function is now vanishing, since $\lim\limits_{t\to\infty} \varphi(t) = 0$ and $\mathbf{1}f(t) = \varphi(t) + A \cdot \mathbf{1}$.

Any non-vanishing disturbance can therefore be represented as the sum of a vanishing disturbance $\varphi(t)$ and the function $A \cdot \mathbf{1}$. Because of the property of superposition* the process caused by the disturbance $\mathbf{1} \cdot f(t)$ can be constructed by adding the ordinates of the process caused by the vanishing disturbance $\varphi(t)$ and the ordinates of the process caused by the unit function $\mathbf{1}$ first increased A times. The construction of the process caused by a vanishing disturbance was described above. To construct the process caused by any non-vanishing disturbance therefore it remains only to explain a method for constructing the process caused by a unit disturbance.

When there is a unit disturbance, on the basis of (4.9) we may write

$$\Phi_x^*(i\omega) = \Phi(i\omega)\,\Phi_{f(t)}(i\omega) = \frac{\Phi(i\omega)}{i\omega},$$

* Since only linear systems are being considered.

since in this case

$$\Phi_{f(t)}(i\omega) = \frac{1}{i\omega}\,.$$

Let

$$\Phi(i\,\omega) = P(\omega) + i\,Q(\omega)\,.$$

Then

$$\Phi_x^*(i\omega) = \frac{Q(\omega)}{\omega} - i\,\frac{P(\omega)}{\omega}\,.$$

It is not possible to use these relations directly in equations (4.16) and (4.17), since the function $\Phi^*(i\,\omega)$ obtained for the unit disturbance tends to ∞ as ω tends to 0. We therefore consider the function

$$\overline{\Phi(i\omega)} = \frac{\Phi(j\omega) - P(0)}{i\omega}\,, \tag{4.21}$$

where $P(0)$ is the value of $P(\omega)$ for $\omega = 0$, and is, of course, constant.

If we could have used formula (4.16) in the given case, then as a result of substituting in it

$$P^*(\omega) = \operatorname{Re}\Phi_x^*(i\,\omega)$$

we would have obtained the required function $x(t)$. If, therefore, we now substitute in (4.16) the real part of the function $\overline{\Phi(i\,\omega)}$ (and this can now be done, since this function is bounded), we obtain not $x(t)$, but another function $x(t)$ which differs from $\overline{x(t)}$ by the function $P(0)\cdot\boldsymbol{1}$:

$$\overline{x(t)} = x(t) - P(0)\cdot\boldsymbol{1}:$$

since the spectrum of the constant $P(0)$ is equal to $\dfrac{P(0)}{i\omega}$

We put

$$\overline{P(\omega)} = \operatorname{Re}\overline{\Phi(i\omega)} = \frac{Q(\omega)}{\omega}\,,$$

$$\overline{Q(\omega)} = \operatorname{Im}\overline{\Phi(i\omega)} = \frac{P(0) - P(\omega)}{\omega}\,.$$

Putting this in equations (4.17) and (4.18) in place of $P^*(\omega)$ and using $Q^*(\omega)$

$$\int_0^\infty \frac{\sin t\omega}{\omega}\,d\omega = \frac{\pi}{2}$$

and

$$x(t) = \overline{x(t)} + P(0),$$

for $t > 0$ we obtain

$$\boxed{x(t) = \frac{2}{\pi}\int_0^\infty \frac{P(\omega)}{\omega}\sin\omega t\,d\omega} \tag{4.22}$$

and

$$\boxed{x(t) = P(0) + \frac{2}{\pi}\int_0^\infty \frac{Q(\omega)}{\omega}\cos\omega t\,d\omega.} \tag{4.23}$$

Thus, when the action on the system is a unit disturbance, the transient process can be expressed as the integral (4.22) or (4.23), where $P(\omega)$ is the real and $Q(\omega)$ is the imaginary part of the frequency characteristic of the given closed system from the point of application of the unit action to the coordinate x. In contrast to the functions $P(\omega)$ and $Q(\omega)$, used in the construction of the process caused by a vanishing disturbance, the functions $P^*(\omega)$ and $Q^*(\omega)$ are determined only by the properties of the system, and do not include the spectrum of the action.

In order to compute this integral we construct the graph of the real frequency characteristic $P(\omega)$ and replace $P(\omega)$ by the algebraic sum of several trapezia with one vertex at the origin of coordinates and two sides coincident with the coordinate axes :

$$P(\omega) = \sum_{j=1}^{k} \lambda_j(\omega), \tag{4.24}$$

exactly as we did above in applying the second method for constructing the process for a vanishing disturbance.

Putting (4.24) in (4.22) we obtain :

$$x(t) = \frac{2}{\pi} \sum \int_0^\infty \lambda(\omega) \frac{\sin \omega t}{\omega} \, d\omega \,, \tag{4.25}$$

where the summation is made over all the trapezia.

Let us now consider in more detail one of the trapezia (Fig. 157) and the integral in the sum (4.25) which corresponds to it namely,

$$x(t) = \frac{2}{\pi} \int_0^\infty \frac{P(\omega)}{\omega} \sin \omega t \, d\omega \tag{4.26}$$

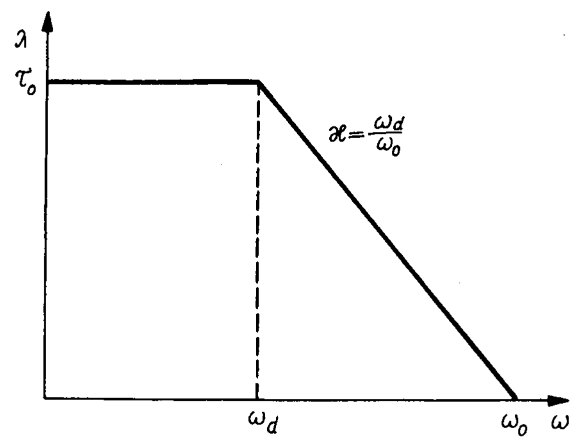


FIG. 157

The trapezium is completely defined by the three numbers τ_0, ω_0 and $\varkappa = \dfrac{\omega_d}{\omega_0}$ (Fig. 157).

The value of the integrals (4.26) can be tabulated and from this table we can determine at once the values of the integral (4.25) for any t.

Such tables would not be very convenient, since for each t they would depend on the three parameters $\varkappa_0$, ω_0 and $\varkappa$.

Because of this, we consider the unit trapezium, which has $\tau_0 = 1$ and $\omega_0 = 1$ (Fig. 158). We denote this unit trapezium by λ_1

and introduce the relation

$$h(t) = \frac{2}{\pi} \int_0^\infty \lambda_1 \frac{\sin \omega t}{\omega} d\omega . \tag{4.27}$$

The integral (4.27) depends for each t on the one parameter $\varkappa$ and it is easy to tabulate its values.

Carrying out the operation of integration in equation (4.27) we obtain :

$$h(t) = \frac{2}{\pi} \left\{ \mathrm{Si}(\varkappa t) + \frac{1}{1-\varkappa} \left[\mathrm{Si}(t) - S(\varkappa t) + \frac{\cos t - \cos \varkappa t}{t} \right] \right\},$$

where

$$\mathrm{Si}\, z = \int_0^z \frac{\sin z}{z} dz .$$

The values of $h(t)$ for $\varkappa$ from 0·00 to 1·00 and for t from 0·00 to 26·0 are set out in Table 15*.

We denote by $\bar{h}(t)$ the value of the integral (4.26) for the given trapezia (for which the conditions $\tau_0 = 1$ $\omega_0 = 1$ do not hold).

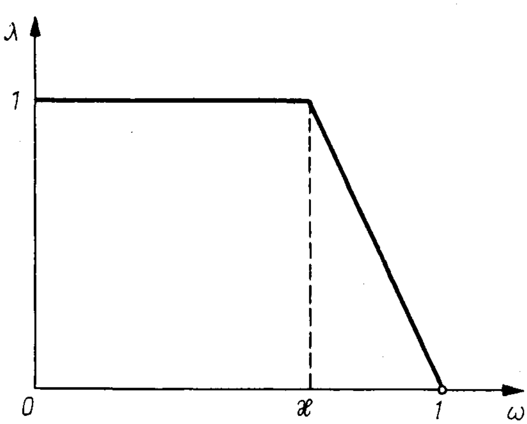


Fig. 158

* There is a much fuller table of the function $h(t)$ in: V. V. Solodovnikov, Yu. I. Topcheyev, G. V. Krutikova, The Frequency Method for the Construction of Transient Processes with the Application of Tables and Nomograms (Chastotnyi metod postroyeniya perekyhodnkh protsessov s prilozheniyem tablits i nomogramm) Gostechizdat (1955).

TABLE

TABLE OF THE

t \ x	0·0	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45
0·0	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000
0·5	0·158	0·166	0·175	0·182	0·190	0·197	0·205	0·213	0·221	0·228
1·0	0·310	0·325	0·341	0·356	0·371	0·386	0·402	0·417	0·432	0·447
1·5	0·449	0·471	0·493	0·515	0·537	0·559	0·580	0·601	0·622	0·642
2·0	0·571	0·600	0·628	0·655	0·682	0·709	0·733	0·761	0·785	0·810
2·5	0·673	0·706	0·739	0·771	0·802	0·832	0·861	0·889	0·916	0·941
3·0	0·755	0·792	0·828	0·863	0·895	0·928	0·958	0·986	0·013	1·038
3·5	0·814	0·854	0·892	0·929	0·963	0·995	1·024	1·051	1·076	1·097
4·0	0·856	0·898	0·937	0·974	1·008	1·038	1·066	1·090	1·110	1·127
4·5	0·882	0·924	0·964	1·000	1·032	1,060	1·084	1·104	1·120	1·131
5·0	0·895	0·939	0·977	1·012	1·042	1·067	1·087	1·102	1·112	1·117
5·5	0·901	0·944	0·982	1·015	1·042	1·063	1·079	1·093	1·092	1·091
6·0	0·903	0·945	0·981	1·013	1·037	1·054	1·065	1·069	1·068	1·062
6·5	0·903	0·945	0·979	1·009	1·030	1·044	1·050	1·050	1·044	1·030
7·0	0·904	0·945	0·978	1·006	1·024	1·034	1·037	1·033	1·023	1·009
7·5	0·906	0·948	0·979	1·005	1·021	1·027	1·027	1·020	1·007	0·991
8·0	0·911	0·951	0·983	1·007	1·020	1·024	1·021	1·011	0·998	0·982
8·5	0·917	0·959	0·989	1·011	1·022	1·024	1·018	1·007	0·993	0·978
9·0	0·925	0·966	0·996	1·016	1·025	1·025	1·017	1·006	0·992	0·978
9·5	0·932	0·973	1·003	1·021	1·028	1·026	1·018	1·005	0·993	0·982
10·0	0·939	0·980	1·009	1·025	1·030	1·027	1·018	1·005	0·994	0·985
10·5	0·944	0·985	1·013	1·028	1·031	1·026	1·016	1·004	0·994	0·988
11·0	0·947	0·988	1·015	1·028	1·030	1·024	1·013	1·002	0·993	0·990
11·5	0·949	0·989	1·015	1·027	1·028	1·020	1·009	0·998	0·992	0·991
12·0	0·950	0·990	1·015	1·025	1·024	1·015	1·004	0·994	0·989	0·990
12·5	0·950	0·990	1·013	1·022	1·019	1·009	0·998	0·988	0·986	0·989
13·0	0·950	0·989	1·012	1·019	1·015	1·004	0·993	0·986	0·984	0·989
13·5	0·950	0·989	1·011	1·016	1·011	1·000	0·990	0·983	0·984	0·989
14,0	0·951	0·990	1·010	1·015	1·008	0,997	0·987	0·983	0·985	0·991
14·5	0·953	0·991	1·011	1·014	1·006	0·995	0·986	0·984	0·987	0·994
15·0	0·956	0·993	1·012	1·014	1·006	0·995	0·987	0·986	0·991	0·998
15·5	0·958	0·996	1·013	1·014	1·006	0·995	0·989	0·989	0·995	1·003
16·0	0·961	0·998	1·015	1·014	1·006	0·995	0·990	0·992	0·999	1·007
16·5	0·963	1·000	1·016	1·015	1·005	0·996	0·992	0·995	1·003	1·010
17·0	0·965	1·001	1·016	1·014	1·005	0·996	0·993	0·998	1·006	1·011
17·5	0·966	1·002	1·016	1·013	1·003	0·995	0·994	0·998	1·007	1·011
18·0	0·966	1·002	1·015	1·012	1·002	0·994	0·994	1·000	1·007	1·010
18·5	0·966	1·002	1·014	1·010	1·000	0·993	0·994	1·001	1·007	1·008
19·0	0·966	1·002	1·013	1·008	0·998	0·992	0·994	1·001	1·006	1·006
19·5	0·967	1·001	1·012	1·006	0·996	0·991	0·994	1·001	1·005	1·003
20·0	0·967	1·001	1·011	1·004	0·995	0·991	0·994	1·001	1·004	1·001

XV
h-FUNCTION

0·50	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00
0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000
0·236	0·244	0·252	0·256	0·265	0·275	0·283	0·294	0·299	0·305	0·313
0·461	0·476	0·491	0·505	0·519	0·534	0·548	0·561	0·575	0·590	0·602
0·662	0·682	0·701	0·720	0·741	0·757	0·775	0·792	0·810	0·827	0·842
0·831	0·856	0·878	0·899	0·919	0·938	0·957	0·974	0·991	1·008	1·022
0·963	0·988	1·010	1·030	1·048	1·066	1·082	1·096	1·109	1·121	1·131
1·061	1·081	1·100	1·116	1·131	1·143	1·154	1·162	1·169	1·174	1·177
1·116	1·133	1·147	1·157	1·165	1·171	1·174	1·175	1·174	1·175	1·166
1·141	1·151	1·158	1·162	1·163	1·161	1·156	1·150	1·141	1·132	1·119
1·138	1·141	1·141	1·138	1·131	1·122	1·111	1·098	1·083	1·069	1·053
1·117	1·114	1·107	1·097	1·084	1·069	1·053	1·036	1·019	1·003	0·987
1·086	1·076	1·064	1·048	1·031	1·014	0·996	0·978	0·963	0·948	0·936
1·051	1·036	1·020	1·001	0·984	0·966	0·949	0·934	0·922	0·914	0·907
1·018	1·001	·0983	0·965	0·948	0·933	0·920	0·911	0·906	0·904	0·906
0·922	0·975	0·957	0·941	0·927	0·917	0·911	0·909	0·911	0·917	0·926
0·975	0·958	0·943	0·931	0·923	0·919	0·920	0·926	0·935	0·946	0·962
0·966	0·952	0·941	0·934	0·932	0·936	0·944	0·955	0·970	0·987	1·002
0·964	0·954	0·948	0·947	0·952	0·961	0·975	0·991	1·008	1·024	1·037
0·968	0·962	0·961	0·967	0·976	0·990	1·006	1·023	1·038	1·051	1·060
0·975	0·973	0·977	0·987	1·000	1·016	1·032	1·047	1·058	1·065	1·066
0·982	0·984	0·993	1·006	1·020	1·036	1·049	1·059	1·063	1·062	1·056
0·988	0·994	1·005	1·019	1·033	1·046	1·054	1·057	1·054	1·046	1·033
0·993	1·001	1·014	1·027	1·039	1·047	1·048	1·044	1·034	1·021	1·005
0·996	1·006	1·018	1·029	1·036	1·038	1·034	1·024	1·010	0·994	0·978
0·997	1·007	1·018	1·026	1·029	1·025	1·015	1·000	0·985	0·970	0·958
0·997	1·007	1·015	1·020	1·017	1·009	0·996	0·979	0·965	0·955	0·950
0·997	1·006	1·012	1·012	1·005	0·993	0·979	0·965	0·955	0·952	0·955
0·998	1·005	1·008	1·004	0·994	0·982	0·968	0·958	0·955	0·960	0·970
0·999	1·005	1·005	0·998	0·987	0·975	0·965	0·961	0·965	0·976	0·991
1·002	1·005	1·002	0·994	0·983	0·974	0·969	0·972	0·982	0·997	1·013
1·005	1·006	1·002	0·993	0·983	0·977	0·978	0·987	1·001	1·018	1·032
1·008	1·007	1·001	0·992	0·985	0·984	0·990	1·003	1·019	1·032	1·039
1·010	1·008	1·001	0·994	0·990	0·993	1·003	1·018	1·031	1·040	1·039
1·012	1·008	1·001	0·995	0·995	1·001	1·014	1·027	1·035	1·037	1·028
1·012	1·007	1·000	0·996	0·999	1·008	1·021	1·030	1·032	1·026	1·012
1·010	1·004	0·998	0·997	1·002	1·012	1·022	1·027	1·022	1·011	0·994
1·008	1·001	0·997	0·997	1·004	1·014	1·020	1·019	1·008	0·993	0·978
1·004	0·998	0·994	0·997	1·005	1·012	1·014	1·007	0·994	0·979	0·969
1·001	0·995	0·993	0·997	1·004	1·009	1·006	0·995	0·981	0·970	0·967
0·997	0·992	0·992	0·997	1·003	1·005	0·998	0·985	0·974	0·969	0·973
0·995	0·991	0·992	0·998	1·003	1·001	0·991	0·980	0·972	0·975	0·986

TABLE XV

TABLE OF THE

t \ x	0·0	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45
20·5	0·967	1·002	1·010	1·003	0·994	0·991	0·995	1·001	1·003	1·000
21·0	0·968	1·002	1·010	1·003	0·994	0·991	0·996	1·002	1·003	0·999
21·5	0·970	1·003	1·010	1·002	0·994	0·993	0·998	1·003	1·003	0·998
22·0	0·971	1·004	1·011	1·002	0·994	0·994	1·000	1·004	1·004	0·998
22·5	0·972	1·005	1·011	1·002	0·995	0·996	1·002	1·006	1·004	0·998
23·0	0·973	1·006	1·011	1·002	0·995	0·997	1·003	1·006	1·004	0·998
23·5	0·974	1·006	1·011	1·002	0·995	0·998	1·004	1·006	1·003	0·998
24·0	0·975	1·006	1·010	1·001	0·995	0·998	1·005	1·006	1·002	0·998
24·5	0·975	1·006	1·009	1·000	0·995	0·999	1·004	1·005	1·000	0·997
25·0	0·975	1·006	1·008	0·999	0·995	0·999	1·004	1·004	0·999	0·996
25·5	0·975	1·006	1·007	0·998	0·994	0·999	1·004	1·002	0·997	0·996
26·0	0·975	1·005	1·006	0·997	0·994	0·999	1·003	1·001	0·996	0·996

If we consider the unit trapezium with the same value of $\varkappa$ as the given trapezium $\tau(\omega)$, then $h(t)$ and $\overline{h(t)}$ are connected by the relation.*

$$\overline{h(t)} = \tau_0 h(\omega_0 t). \tag{4.28}$$

The product $\omega_0 t$ is dimensionless. In Table 15 the corresponding term is shown as dimensionless time. For $\omega_0 = 1$ it coincides numerically with t in seconds.

The relation (4.28) enables us to determine the value of $\overline{h(t)}$ for any moment t from the table constructed for the unit trapezium.

From (4.25) the required transient process can be found from the formula $x(t) = \Sigma\, h(t)$, where the summation is made over all the trapezia.

(d) *The construction of the real frequency characteristic $P(\omega)$*

When these methods for constructing the control process are used in practice, the greatest time is spent in constructing the real frequency characteristic $P(\omega)$ for the given system. If, during the analysis of stability, the D-partition was constructed for the total

* See Section 7, Criterion 10 (p. 323).

(contd.)

h-FUNCTION

0·50	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·85	1·00
0·994	0·991	0·994	1·000	1·002	0·998	0·987	0·978	0·977	0·987	1·001
0·993	0·992	0·996	1·001	1·002	0·996	0·986	0·982	0·987	1·001	1·015
0·994	0·994	0·999	1·004	1·002	0·995	0·988	0·988	0·998	1·014	1·025
0·995	0·997	1·002	1·005	1·002	0·995	0·992	0·997	1·010	1·024	1·029
0·996	1·000	1·005	1·007	1·002	0·996	0·996	1·005	1·019	1·028	1·028
0·997	1·002	1·007	1·007	1·002	0·997	1·001	1·011	1·022	1·025	1·016
0·999	1·003	1·008	1·006	1·001	0·999	1·004	[illegible]·015	1·021	1·016	1·003
0·999	1·004	1·007	1·004	0·999	0·999	1·007	1·015	1·016	1·006	0·990
0·999	1·004	1·006	1·002	0·997	0·999	1·007	1·012	1·007	0·993	0·980
1·000	1·004	1·004	0·999	0·996	1·000	1·007	1·008	0·998	0·984	0·975
1·000	1·003	1·001	0·996	0·995	1·000	1·005	1·001	0·989	0·978	0·977
1·000	1·002	0·999	0·995	0·995	1·000	1·002	0·996	0·984	0·978	0·984

coefficient of amplification of the open system, then this construction can be used to simplify considerably the construction of the real characteristic $P(\omega)$.

Let the transfer function from the point of application of the disturbance to the given coordinate be of the form

$$\Phi(p) = \frac{L(p)}{D(p) + KR(p)}, \tag{4.29}$$

where K is the total coefficient of amplification of the open circuit. Dividing the numerator and denominator of $\Phi(i\omega)$ by $R(i\omega)$ we obtain :

$$\Phi(i\omega) = \frac{\dfrac{L(i\omega)}{R(i\omega)}}{\dfrac{D(i\omega)}{R(i\omega)} + K}.$$

The denominator of $\Phi(p)$ is the same as the left-hand side of the characteristic equation of the system. Hence, the boundary of the D-partition for K is :

$$K = S(i\omega) = -\frac{D(i\omega)}{R(i\omega)}.$$

Putting this value in $\Phi\,(i\,\omega)$, we find :

$$\Phi\,(i\omega) = \frac{-\dfrac{L\,(i\omega)}{R\,(i\omega)}}{S\,(i\omega) - K}\,.$$

We now add the hodograph of the vector $-\dfrac{L\,(i\omega)}{R\,(i\omega)}$ to the graph on which the D-partition boundary for K (i. e. the hodograph of the vector $S\,(i\,\omega)$ was drawn. Then $\Phi\,(i\,\omega)$ is defined as the ratio of the vectors $-\dfrac{L(i\omega)}{R\,(i\omega)}$ and $S\,(i\,\omega) - K$, taken for the same value of ω, and the real characteristic $P\,(\omega)$ is equal to

$$P\,(\omega) = \frac{A}{B}\cos\,(\varphi_1 - \varphi_2)\,,$$

where A and B are the moduli and φ_1 and φ_2 are the arguments of the vectors $-\dfrac{L\,(i\omega)}{R\,(i\omega)}$ and $S\,(i\,\omega) - K$ for the same value of ω. One end of the vector $S\,(i\,\omega\,) - K$ lies at the point $u = K$ and the others at the point on the hodograph of $S\,(i\,\omega)$ corresponding to this value of ω. When $S\,(i\,\omega)$ has already been constructed for the analysis of stability, the calculation of $P\,(\omega)$ reduces to the additional construction of the points of the hodograph of $-\dfrac{L\,(i\omega)}{R\,(i\omega)}$, which is usually not very complicated since the degree of $L(p)$ and of $R(p)$ is not high. In servo-mechanism theory and in some problems in the theory of automatic control the disturbances which are applied are such that the transfer function of the closed system is

$$\Phi\,(p) = \frac{W\,(p)}{1 + W(p)}\,,$$

where $W\,(p)$ is the transfer function of the open system. In this case

$$\Phi\,(i\omega) = \frac{K}{-\,S\,(i\omega) + K}\,.$$

If from any point ω_j of the hodograph $S\,(i\,\omega)$ a vector is drawn to the

point a where $u = K$ (Fig. 159), then this vector determines the denominator in (4.29).

Therefore

$$P(\omega_j) = \frac{\overline{Oa}}{\overline{ab}} \cos \varphi = \frac{\overline{ac}}{\overline{ab}} .$$

In this particular case, the real frequency characteristic $P(\omega)$ can be constructed immediately if we know the boundary of the D-partition for K. In order to do this a perpendicular is dropped from

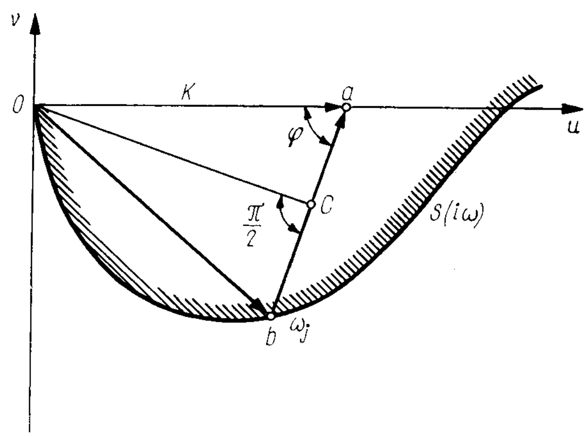


Fig. 159

the origin of coordinates to the vector $\overline{ab}$, corresponding to any frequency ω_j, and we take the ratio of the segment ac cut off on $\overline{ab}$ to the magnitude of this vector (Fig. 159).

Doing this for various frequencies, we can construct the points of the whole real frequency characteristic.

In just the same way it is easy to find a method for calculating $P(\omega)$ from the amplitude-phase characteristic of the open system of the first kind $\frac{D(i\omega)}{K(i\omega)}$ or of the second kind $\frac{K(i\omega)}{D(i\omega)}$ *.

* Another method for constructing $P(\omega)$ from the amplitude-phase characteristic of the open system is described in the book mentioned in the footnote on p. 283.

5. General Considerations of Indirect Estimates of the Control Process. The Degree of Stability

Each of the methods described above enables us, with greater or less difficulty and expenditure of time, to construct the control process for a given disturbance and for fixed values of all the parameters. Methods of constructing the process which do not involve the calculation of the roots of the characteristic equation make it somewhat easier to determine how the assembly parameters influence the course of the process than do methods which require the calculation of these roots, but in either case it is a complex and laborious problem to find the dependence of the processes on the change in the parameters. Also, the technical conditions laid on the design of controllers usually place strict requirements on the character of the process, and the problem of designing the load and the tuning of these devices consists of satisfying these requirements. In such cases the construction of the process is used only as a check, and the choice of parameters is made according to these or other indirect estimates. We shall restrict ourselves henceforth to the consideration of estimates of transient processes, i.e. we shall consider the disturbance to be the unit function. In the theory of automatic control three types of indirect estimate of the transient process are used : estimates concerning the distribution of zeroes (or zeroes and poles) of the transfer function $\Phi(p)$, the integrals of the estimate, and estimates of the process from the graphs of the frequency characteristics of the given system.

Among the estimates concerning the distribution of the zeroes and poles of the transfer function, the simplest and most widely used is that which finds the distance in the root-plane between the imaginary axis and the root nearest to it. This distance is called the "*degree of stability*".

(a) *The concept of the degree of stability*

We return to the formula which defined the control process in terms of the roots of the characteristic equation

$$x(t) = \sum_{k=1}^{n} \frac{A(p_k)}{B'(p_k)} e^{p_k t} = \sum_{k=1}^{n} c_k e^{p_k t},$$

where p_k is a root of the characteristic equation.

Those terms for which the real part of the root $p_k = -\alpha_k \pm \pm i\beta_k$ is small in modulus, i.e. for which the root lies near the imaginary axis, dominate this sum. In such cases we sometimes estimate the transient process by finding the distance between the imaginary axis and the root nearest to it, i.e. by finding the degree of stability. Of course, the system is here assumed to be stable.

Let us denote the degree of stability by δ.

The degree of stability is said to be *aperiodic* if the root nearest the imaginary axis is entirely real. It is said to be *oscillatory* if the root nearest the imaginary axis is complex. If the degree of stability is aperiodic, then the exponential determined by this root will dominate the sum. The damping of the transient process is determined roughly by this exponential.

If the degree of stability is oscillatory, then the process is characterized roughly by the damped sinusoid, whose envelope has the equation

$$x = Ce^{-\delta t}.$$

Therefore the value of δ is sometimes indirectly measured by the damping of the process.

(b) *The determination of the magnitude of the degree of stability from the Routh-Hurwitz criterion*

Let the characteristic equation of the system be given:

$$p^n + a_1 p^{n-1} + a_2 p^{n-2} + \ldots + a_{n-1} p + a_n = 0.$$

We reduce this equation to a simpler one, by making the substitution

$$p = \sqrt[n]{a_n}\,\lambda.$$

Then

$$f(\lambda) = \lambda^n + A_1 \lambda^{n-1} + A_2 \lambda^{n-2} + \ldots + A_{n-1}\lambda + 1 = 0,$$

where

$$A_k = \frac{a_k}{\sqrt[n]{a_n^k}}.$$

We shift the imaginary axis to the left so that it passes through the root nearest to it. In order to do this we make the substitution $\lambda = z - \delta$:

$$f(z-\delta) = (z-\delta)^n + A_1 (z-\delta)^{n-1} + \\ + A_2 (z-\delta)^{n-2} + \ldots + A_{n-1}(z-\delta) + 1 = 0 .$$

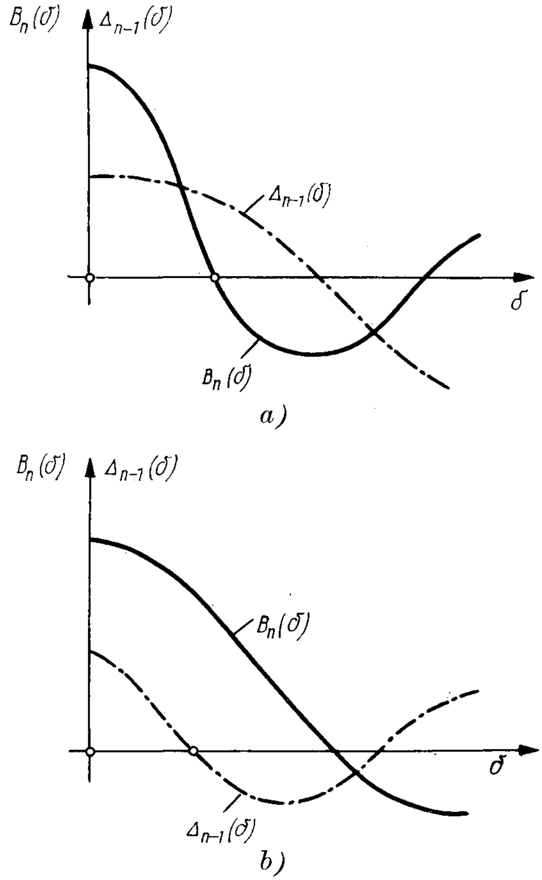


FIG. 160

Opening the brackets and collecting like terms, we obtain:

$$f(z) = z^n + B_1 z^{n-1} + B_2 z^{n-2} + \ldots + B_{n-1} z + B_n = 0 ,$$

where

$$B_k = \frac{1}{(n-k)!} \left[\frac{\mathrm{d}^{n-f}(z)}{\mathrm{d}z^{n-k}} \right]_{z=-\delta} .$$

For $\delta = 0$ all roots of the equation $f(z) = 0$ lie to the left of the imaginary axis, since $f(z)$ is then equal to $f(\lambda)$, but we have assumed the system to be stable. All the Hurwitz determinants $\Delta_1, \Delta_2, \ldots, \Delta_n$, therefore, formed from the coefficients of the equation $f(z)$ for $\delta = 0$, are positive. We now increase δ until one of the Hurwitz determinants first becomes zero. The corresponding value of δ will be the required degree of stability. As δ increases the highest determinant Δ_n is always the first to become zero. But $\Delta_n = B_n \Delta_{n-1}$ and Δ_n can only

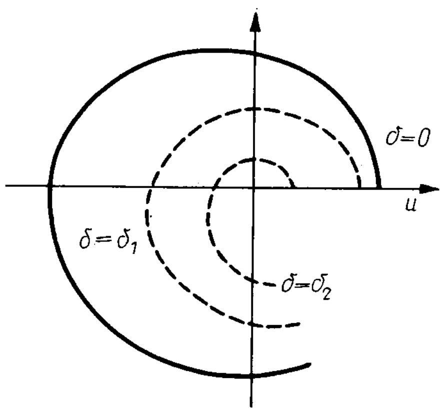


FIG. 161

become zero either because $B_n = 0$ or because $\Delta_{n-1} = 0$. The degree of stability is aperiodic if in this case $B_n = 0$, and is oscillatory if $\Delta_{n-1} = 0$. This leads to the following simple method for determining the degree of stability. We construct the curves $B_n(\delta)$ and $\Delta_{n-1}(\delta)$ and find the point corresponding to the intersection of one of these curves with the δ-axis which is nearest the origin of co-ordinates (Fig. 160). This point determines δ, the degree of stability. It is aperiodic when (Fig. 160*a*) this point belongs to the curve $B_n(\delta)$ and it is oscillatory when (Fig. 160*b*) it belongs to the curve $\Delta_{n-1}(\delta)$.

(c) *The determination of the degree of stability from the Mikhailov criterion*

In order to determine the degree of stability we can use any of the stability criteria. We have just considered a method based on the use of the Routh-Hurwitz criterion. We now determine the degree of stability using the Mikhailov criterion.

Let the characteristic equation of the given system be reduced to

$$z^n + B_1 z^{n-1} + B_2 z^{n-2} + \ldots + B_{n-1} z + B_n = 0.$$

We draw its Mikhailov hodograph, first taking $\delta = 0$. In this case,

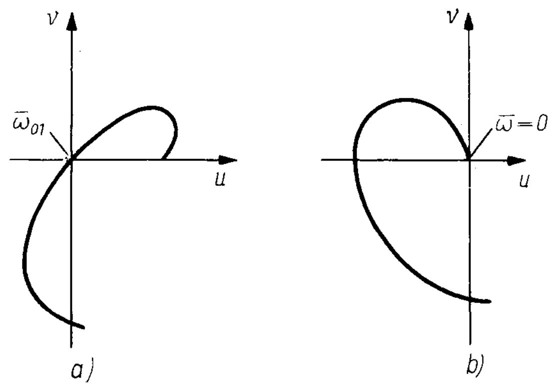


FIG. 162

by hypothesis, the system is stable. We vary the parameter δ, giving to it various values, and for each of them we shall construct its corresponding Mikhailov hodograph (Fig. 161). We continue this operation

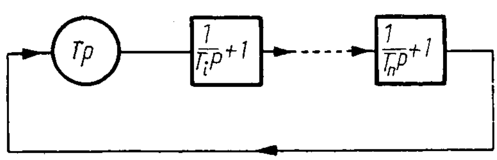


FIG. 163

until for some δ the hodograph first passes through the origin of coordinates. This value of δ is equal to the required degree of stability. The Mikhailov hodograph can pass through the origin in different ways. Two cases are possible (Fig. 162*a* and *b*).

In the first case, the degree of stability will be oscillatory, and the imaginary part of the root occurring on the imaginary axis is found at once to be the point corresponding to $\overline{\omega_{01}}$. In the second case the degree of stability will be aperiodic, since in this case the free term beomes zero : $B_n = 0$.

To determine the degree of stability we need only construct the small part of the Mikhailov curve near the origin of coordinates

The convenience of using the degree of stability as an indirect estimate of the process arises from the fact that in the plane of any one or two parameters which enter linearly into the characteristic equation it is easy to construct the curves determining the lines corresponding to the degree of stability. To do this it is only necessary to know the value of δ and then to construct the boundary of the region of stability, for example, by the usual construction of the D-partition for the equation $f(z) = 0$. In this way, by constructing the

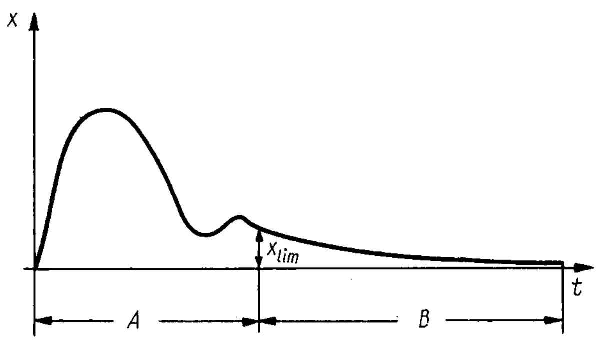


FIG. 164

D-partition for the equation $f(z) = 0$ several times, we can not only determine the region of stability (putting $\delta = 0$) but also partition it by lines equal to δ.

(d) *The connexion between the degree of stability and the values of the time constants of the stages of a single-loop circuit*

When considering questions concerning the stability of single-loop systems (without derivative action) we established that these systems can be made stable only by choosing the parameters of the stages to fulfil definite conditions, concerning the structure of the network itself: the system cannot be made stable if it contains more than one stage which is astatic or unstable, or if it contains so many conservative stages that the inequality $n > 4r$ is not satisfied, where n is the degree of the characteristic equation, and r is the number of conservative stages.

Let us first consider a system containing one astatic stage and any number of single-capacitance stages.

The structural scheme is similar to the system shown in Fig. 163. The characteristic equation of the system is

$$p\,(T_1 p + 1)\,(T_2 p + 1) \ldots (T_n p + 1) + K = 0.$$

We make the substitution $p = z - \delta$, obtaining

$$(z - \delta)\,(T_1 z + 1 - \delta\, T_1)\,(T_2 z + 1 - \delta\, T_2) \ldots$$
$$\ldots (T_n z + 1 - \delta\, T_n) + K = 0.$$

The degree of stability of the given system is greater than any value $\delta = \delta_0$, if, after putting $\delta = \delta_0$ in this equation, all its roots lie to the left of the imaginary axis. To be precise let us suppose that the stages of the system are so numbered that the values of the time constants decrease, i.e.

$$T_1 > T_2 > T_3 > \ldots > T_n\,.$$

Depending on the increase in δ, the value $\varrho = \dfrac{1}{T_1}$ will be reached at some point. For large δ at least two brackets will have negative free terms, and this corresponds to a system containing more than one unstable stage. Such a system is structurally unstable, i.e. for any choice of the parameters T_i $(i = 1, 2, \ldots, n)$ and K not all the roots of the given equation can lie to the left of the imaginary axis. Hence, the degree of stability of the given system cannot be greater than $\dfrac{1}{T_1}$, i.e. *the degree of stability does not exceed the reciprocal of the greatest time constant of the single-capacitance stages.*

We now consider a single-loop system consisting only of single-capacitance stages. The characteristic equation of this system can be written in the following form :

$$\prod_{j=1}^{n} (T_j\, p + 1) + K = 0\,.$$

We make the substitution $p = z - \delta$:

$$\prod_{j=1}^{n} [T_j z + (1 - T_j\, \delta)] + K = 0.$$

For $\delta = 0$ we obtain the initial equation. As δ increases, the stage with the greatest time constant will become unstable, but its presence can still not destroy the structural stability of the system. As δ increases further, the stage with the second largest time constant becomes unstable, and from this value of δ the structural stability will be lost.

Thus, in this case *the degree of stability, δ, cannot be made greater than $\frac{1}{T_2}$ where T_2 is the second largest time constant in the circuit of single-capacitance stages.* In real systems the second largest time constant is usually considerably smaller than the first, and therefore when there is no astatic stage a considerably larger value of δ can be obtained.

In exactly the same way, by using the theorems of structural stability, we can find the maximum degree of stability attainable in systems containing stages of other types (oscillatory, unstable, conservative).

(e) *Remarks*

The exceptional simplicity of estimating the "degree of stability" and the relative ease with which one may determine the parameters which minimize it often lead to excessive optimization and careless use of this estimate in the design of the control system.

In fact, the use of the degree of stability in the calculation demands great care. Only for very special initial conditions can we succeed in finding a direct connexion between the magnitude of the degree of stability and the quality of the process. Even in these cases the degree of stability is related to only one indicator characterizing the quality of the process, the control time. Usually, if we construct a line representing the degree of stability in the plane of any two of the parameters and compare the processes caused by the same disturbance in systems corresponding to the various points of this line, then these processes are seen to differ very much. Not only the overshoot, but also the control time, can change by several times when the same degree of stability exists for various ratios of the parameters.

Moreover, this estimate is often not generally applicable to a proportional rate floating controller, for the following reason.

The processes in this controller can be divided into two sections A and B (Fig. 164). The section A corresponds to the steady value

of x with an accuracy depending on the static error, determined by the static part of the controller, and the section B to the removal of this static error by the float. Usually the length of the section B is considerably greater than the length of A. But we are interested mostly not in the total time corresponding to the sections A and B but in the time corresponding to the section A. Even when the degree of stability is related to the control time, it determines the total time of the process. The length of A is related in such systems to the distance between the imaginary axis and the root nearest but one to it.

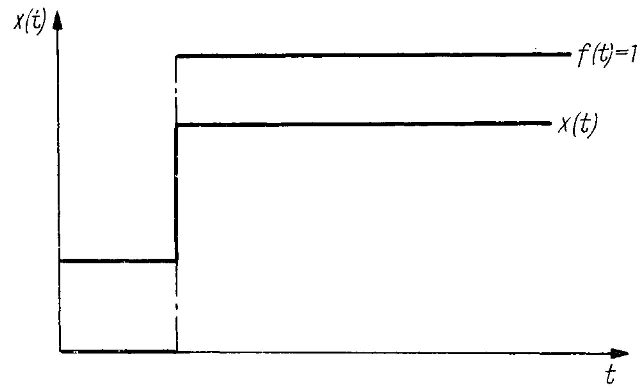


FIG. 165

The question of when we can and when we cannot determine the optimal parameters minimizing the degree of stability can be solved only by experimental investigation and adjustment of real systems. For this reason in control theory we do not limit ourselves to this more simply calculated, and also very primitive, estimate, but we consider more complex indirect estimates.

6. Integral Estimates

(a) *General considerations*

Until now we have everywhere taken the value of the coordinates in the given steady state as the zero of the increments in the coordinates, i.e. we have taken the "old equilibrium" which existed in the system up to the beginning of the disturbing action causing the

control process as the origin. In this section we shall measure the increments of the coordinates from their new equilibrium value, which is set up in the system as a result of the disturbance 1. This restriction is not essential when integral estimates are used. For, by

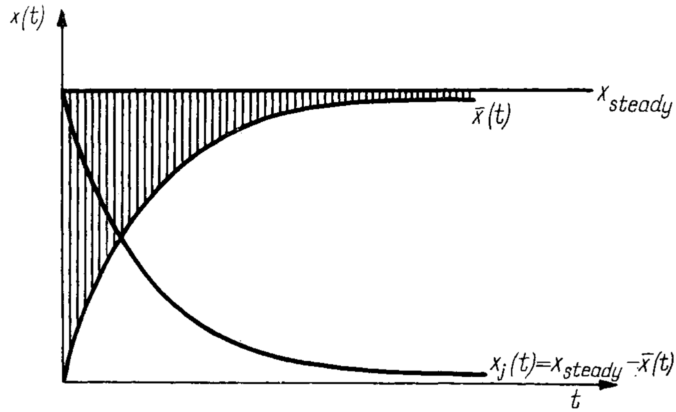


FIG. 166

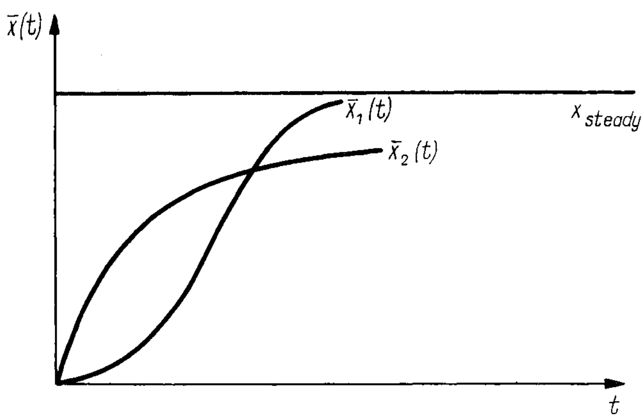


FIG. 167

making the computations more complicated, we can easily extend them to the case when the given disturbances are other functions. Limited here to an estimate of the transient processes we will replace the unit disturbance 1 by the equivalent initial conditions.

The transient process would be ideal if at the moment the unit disturbance appeared, the considered coordinate instaneously

acquired its new steady value and did not change it again until the appearance of a new disturbance (Fig. 165). In real systems such a process is not possible, but the smaller the area between the curves of the real and ideal processes (shaded in Fig. 166), the less the process will differ from the ideal. In the case when there is no overshoot (if the system is static, see Fig. 167) or when the curve $x(t)$ does not intersect the t-axis a second time (if the system is astatic) the given

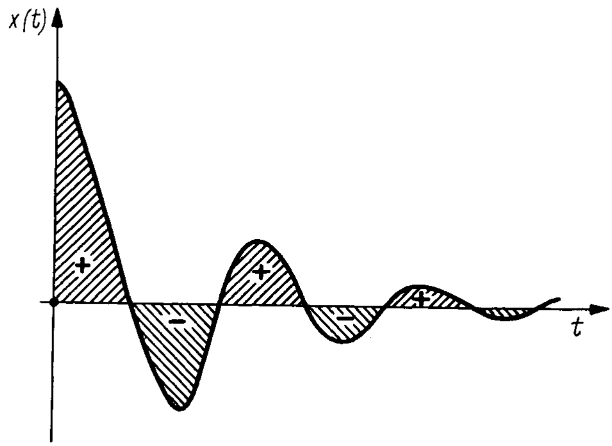


FIG. 168

area is defined by the integral

$$\int_0^\infty x(t)\,\mathrm{d}t\,. \tag{4.30}$$

In other cases this integral does not define the given area, since in computing the integral the separate areas must be summed taking the sign of the ordinates into account (Fig. 168) ; thus, for example, in the case of weakly-damped oscillations independent of the amplitude, the integral would be small, although the area, characterizing the deviation of the real process from the ideal, can here be as large as we please.

In the cases given above, when the integral (4.30) does define the given area, it provides a convenient means of choosing the system parameters : the parameters are chosen so that the value of the integral will be a minimum. Of course, such an estimate is indirect,

and the selection of the parameters must be made as a preliminary, since very different processes can occur even for monotonic processes with equal areas. Despite this, the use of such estimates often enables us to select the initial parameters quickly, verification that the choice is correct follows by constructing the process.

To calculate the integral (4.30) we note that the Laplace transform for the expression $x_j(t)$ is, by definition, equal to

$$L[x_j(t)] = \int_0^\infty x_j(t)\, e^{-pt}\, \mathrm{d}t$$

and hence

$$\int_0^\infty x_j(t)\, \mathrm{d}t = \lim_{p\to 0} L[x(t)].$$

The practical application of such a simple estimate is complicated by the fact that only in rare cases is it known beforehand that the process does not possess overshoot or that, in the astatic case, during the process a zero value of the deviation of the controlled coordinate is not attained several times.

If the process is oscillatory, then an estimate of how close the transient process is to the ideal can be made from the values of the integral $\int_0^\infty |x(t)|\, \mathrm{d}t$, but the value of an integral such as this is usually difficult to compute. It is more convenient to estimate the process from the value of the integral

$$\int_0^\infty x^2(t)\, \mathrm{d}t. \tag{4.31}$$

If we select the system parameters to make this integral a minimum, then the transient process obtained is usually excessively oscillatory.

In order to avoid obtaining strongly oscillating processes, we assume that the parameters have been chosen so that the integral

$$\int_0^\infty [x^2(t) + \tau^2 \dot{x}^2(t)]\, \mathrm{d}t, \tag{4.32}$$

shall be a minimum, where τ is any given number.

The selection of the system parameters made so that we obtain the minimum of this integral (provided we have chosen the right

value of τ) enables us to obtain a sufficiently good transient process with small overshoot and it is often monotonic. Sometimes we use even more complicated estimates

$$\int_0^\infty [x^2(t) + \tau_1^2 \dot{x}^2(t) + \tau_2^4 \ddot{x}^2(t)]\, \mathrm{d}t$$

or, more generally :

$$\int_0^\infty \left\{ x^2(t) + \tau_1^2 \left[\frac{\mathrm{d}x(t)}{\mathrm{d}t}\right]^2 + \tau_4^2 \left[\frac{\mathrm{d}^2 x(t)}{\mathrm{d}t^2}\right]^2 + \ldots + \tau_n^{2n} \left[\frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}\right]^2 \right\} \mathrm{d}t.$$

We will restrict ourselves to the basic estimate (4.32). To use this integral estimate to calculate the system, we must answer the following questions :

1. How must the value of τ, entering in (4.32), be chosen for the investigation of the real automatic control system?

2. How do we find the system parameters for which the chosen estimate is a minimum?

3. How near to the process required by the technical conditions will be the transient process with parameters chosen in this way?

(b) *The choice of the integral estimate*

We write the integral

$$I = \int_0^\infty [x^2(t) + \tau^2 \dot{x}^2(t)]\, \mathrm{d}t$$

as the difference of two integrals :

$$\int_0^\infty [x^2(t) + \tau^2 \dot{x}^2(t)]\, \mathrm{d}t =$$

$$= \int_0^\infty [x(t) + \tau \dot{x}(t)^2\, \mathrm{d}t - 2\tau \int_0^\infty x(t)\, \dot{x}(t)\, \mathrm{d}t =$$

$$= \int_0^\infty [x(t) + \tau \dot{x}(t)]^2\, \mathrm{d}t - 2\tau \int_0^\infty x(t) \frac{\mathrm{d}x}{\mathrm{d}t}\, \mathrm{d}t =$$

$$= \int_0^\infty [x(t) + \tau \dot{x}(t)]^2\, \mathrm{d}t - 2\tau \int_0^\infty x(t)\, \mathrm{d}x.$$

We compute the last integral:

$$2\tau \int_0^\infty x(t)\,\mathrm{d}x = 2\tau \left.\frac{x^2(t)}{2}\right|_0^\infty = \tau\left[x^2(\infty) - x^2(0)\right].$$

If the system is stable, then $x(\infty) = 0$, since in this section we take the origin of x as the value x_{steady} which is set up as $t \to \infty$.

Then

$$\int_0^\infty [x^2(t) + \tau^2 \dot{x}^2(t)]\,\mathrm{d}t = \int_0^\infty [x(t) + \tau\dot{x}(t)]^2\,\mathrm{d}t + \tau x^2(0).$$

The last term on the right hand side is a constant, defined by the initial deviation of the system. The first integral

$$\int_0^\infty [x^2(t) + \tau^2 \dot{x}^2(t)]\,\mathrm{d}t$$

will have its least value when the integral on the right hand side of the equation written above is zero,

$$\int_0^\infty [x(t) + \tau\dot{x}(t)]^2\,\mathrm{d}t = 0.$$

Due to the fact that the integrand is always positive, this can happen only when the integrand is equal to zero, i.e.

$$[x(t) + \tau\dot{x}(t)]^2 = 0$$

or

$$x(t) + \tau\dot{x}(t) = 0. \tag{4.33}$$

Thus, the first integral attains its minimum when $x(t)$ satisfies the differential equation (4.33). The least value $I_{\min\min}$ is equal to

$$I_{\min\min} = \tau x^2(0).$$

The differential equation (4.33) determines the transient process to which we can approximate in the limit, if the parameters can be

chosen so that $I = I_{\min\min}$. This limit process will be defined by the exponential $x(t) = x(0)e^{-\frac{t}{\tau}}$.

The value of τ *is chosen so that the exponential* $x(t) = x(0)e^{-\frac{t}{\tau}}$ *will with a known tolerance satisfy the conditions which must be satisfied by the transient process.* Once τ has been chosen the integral estimate is fixed. After this the system parameters are selected so that the chosen integral estimate takes its least value, $I_{\min}$. Of course, the

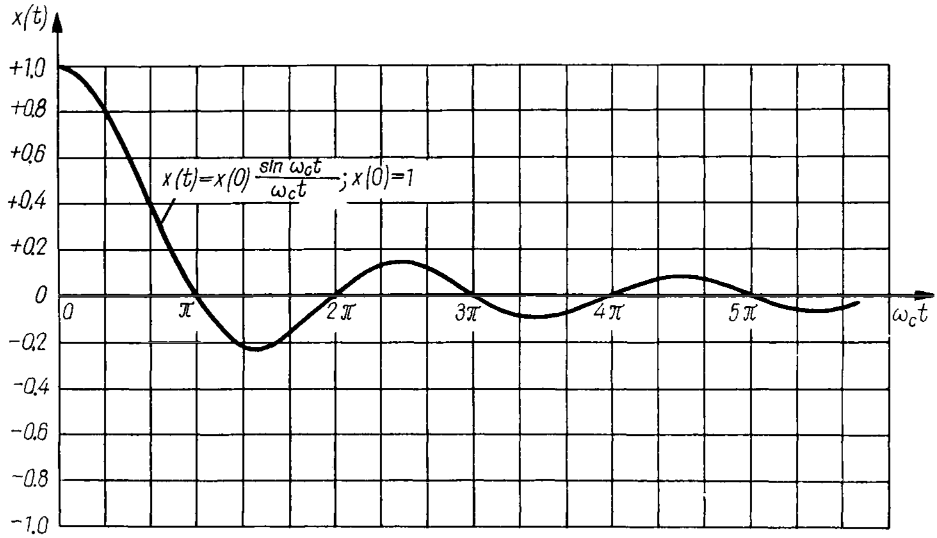


FIG. 169

value of $I_{\min}$ in this real case will be greater than $I_{\min\min}$, and the process in the system will differ from the given exponential. But of all the possible parameters those chosen will ensure the process which is nearest to this exponential.

It has been shown that when the simplest quadratic integral estimate $\int_0^\infty x^2(t)\,dt$ is minimized, the process nearest to

$$x(t) = x(0)\frac{\sin \omega_c t}{\omega_c t}.$$

is ensured. The graph of this function is shown in Fig. 169. It is natural, therefore, that from the point of view of the minimization of the integral, the choice of parameters leads to a system with an excessively oscillatory process.

By using more complicated estimates we can, by minimizing them, approximate to processes of a more complicated form, for example to processes consisting of several summed exponentials.

We pass now to the computation of $\int_0^\infty (x^2 + \tau^2 \dot{x}^2)\,\mathrm{d}t$ via the system parameters and to the choice of the parameters which minimize this integral.

(c) *The determination of the system parameters which minimize the integral estimate*

We consider a system of linear differential equations of the general form*

$$\dot{x}_i = \sum_{j=1}^{n} a_{ij} x_j$$

and the most general form of the integral quadratic estimate

$$I = \int_0^\infty V\,\mathrm{d}t\,,$$

where

$$V = A_1 x_1^2 + A_2 x_2^2 + \ldots + A_n x_n^2 = \sum_{i=1}^{n} A_i x_i^2 .$$

The integral (4.32) which interests us is obtained from this if we put

$$x_1 = x,\ x_2 = \dot{x},\ A_1 = 1,\ A_2 = \tau^2 \quad \text{and} \quad A_3 = A_4 = \ldots = A_n = 0 .$$

We choose another quadratic function U such that

$$\frac{\mathrm{d}U}{\mathrm{d}t} = -V . \tag{4.34}$$

Then the integral $I = \int_0^\infty V\mathrm{d}t$ is easily calculated.

* If the system contains equations whose orders are higher than the first, then it is easy to reduce to this form, by denoting the second and higher derivatives by new letters. Thus, for example, the equation

$$a\ddot{x}_1 + b\dot{x}_1 + cx_1 = 0$$

reduces to the system

$$\left.\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -\frac{b}{a}x_2 - \frac{c}{a}x_1 . \end{aligned}\right\}$$

In fact

$$V\,\mathrm{d}t = -\mathrm{d}U\,,$$

i. e.

$$I = \int_0^\infty V\,\mathrm{d}t = -\int_0^\infty \mathrm{d}U = -\,U\Big|_0^\infty = -\,[U\,(\infty), -\,U\,(0)]\,.$$

In a stable system when $t = \infty$ $x_1 = x_2 \ldots = x_n = 0$, and therefore $U(\infty) = 0$.

Therefore

$$I = \int_0^\infty V\,\mathrm{d}t = U\,(0)\,,$$

i.e. in order to determine I we put the initial values of $x_1, x_2, \ldots, x_n$ in U.

In order to determine U, we will look for it in the form

$$U = \sum_{i,j=1}^{n} B_{ij}\,x_i\,x_j,$$

where all the B_{ij} are numbers which have to be determined so that the equation

$$\frac{\mathrm{d}U}{\mathrm{d}t} = -\,V$$

or

$$\sum_{i=1}^{n} \frac{\partial U}{\partial x_i}\,\dot{x}_i = -\,V\,.$$

is satisfied.

Putting in this the above expressions for V and U we obtain:

$$\sum_{i=1}^{n}\left[\sum_{j=1}^{n} B_{ij}\,x_j\right]\dot{x}_i = -\sum_{i=1}^{n} A_i x_i^2\,.$$

We replace $\dot{x}_i$ at once by the right hand side of the initial system of first order linear equations

$$\sum_{i=1}^{n}\left[\sum_{j=1}^{n} B_{ij}\,x_j\right]\left[\sum_{j=1}^{n} a_{ij}\,x_j\right] = -\sum_{i=1}^{n} A_i\,x_i^2\,.$$

The left and right-hand sides of this equation consist of quadratic forms. Equating coefficients of $x_1^2, x_2^2, \ldots, x_n^2$ on the left and right and equating the coefficients of the products $x_i x_j$ $(i \neq j)$ to zero (since there are no similar terms on the right-hand side) we obtain a system of linear algebraic equations for all the B_{ij}. Solving these equations, substituting the values of B_{ij} in U and using the given initial conditions, we find U as a function, which we can minimize further, of the system parameters (the coefficients a_{ij}).

EXAMPLE 1. We demonstrate the method for finding U by taking as an example the equation

$$a_0 \ddot{x}_1 + a_1 \dot{x}_1 + a_2 x_1 = 0,$$

which we replace by the system

$$\left.\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -\frac{a_1}{a_0} x_2 - \frac{x_2}{a_0} x_1. \end{aligned}\right\}$$

Let us put

$$V = x_1^2 + \tau^2 x_2^2 .$$

We seek U in the form

$$U = B_{x_1} x_1^2 + B_{x_1 x_2} x_1 x_2 + B_{x_2} x_2^2 .$$

From the condition

$$\frac{dU}{dt} = -V$$

we find

$$\frac{\partial U}{\partial x_1} \dot{x}_1 + \frac{\partial U}{\partial x_2} \dot{x}_2 = (2B_{x_1} x_1 + B_{x_1 x_2} x_2)\, x_2 + $$
$$+ (2B_{x_2} x_2 + B_{x_1 x_2} x_1) \left(-\frac{a_1}{a_0} x_2 - \frac{a_2}{a_0} x_1\right) = -x_1^2 - \tau^2 x_2^2 .$$

Equating coefficients on the left and right, we obtain the system of equation

$$\left.\begin{aligned} -\frac{a_2}{a_0} B_{x_1 x_2} &= -1, \\ B_{x_1 x_1} - 2\frac{a_1}{a_0} B_{x_2} &= -\tau^2, \\ 2B_{x_1} - \frac{a_1}{a_2} B_{x_1 x_2} - 2\frac{a_2}{a_0} B_{x_2} &= 0. \end{aligned}\right\}$$

Solving this system, we find

$$B_{x_1x_2} = \frac{a_0}{a_2},$$

$$B_{x_1} = \frac{a_0^2}{2a_1 a_2}\left[\frac{a_2^2}{a_0^2}\tau^2 + \left(\frac{a_1^2}{a_0^2} + \frac{a_2}{a_0}\right)\right],$$

$$B_{x_2} = \frac{a_0}{2a_1}\left[\frac{a_0}{a_2} + \tau^2\right],$$

whence

$$I = \int_0^\infty (x_1^2 + \tau^2 \dot{x}_1^2)\,\mathrm{d}t = U(0) =$$

$$= \frac{a_0^2}{2a_1 a_2}\left[\frac{a_2^2}{a_0^2}\tau^2 + \left(\frac{a_1^2}{a_0^2} + \frac{a_2}{a_0}\right)\right] x_1^2(0) +$$

$$+ \frac{a_0}{2a_1}\left[\frac{a_0}{a_2} + \tau^2\right] x_2^2(0) + \frac{a_0}{a_2} x_1(0)\, x_2(0)\,.$$

The parameters a_0, a_1 and a_2 which minimize I are now easy to find by the usual rules for finding the absolute minimum of a function of several variables.

Example 2. Let the given system be described by the second order equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + h\frac{\mathrm{d}x}{\mathrm{d}t} + 3x = 0\,.$$

Let us suppose that for

$$t = 0 \quad x = 1, \frac{\mathrm{d}x}{\mathrm{d}t} = 0.$$

The integral estimate is $I = \int_0^\infty (x^2 + 2\dot{x}^2)\,\mathrm{d}t$. In the given case

$$x_1 = x,\, x_2 = \frac{\mathrm{d}x}{\mathrm{d}t},\, x_1(0) = 1,\, x_2(0) = 0,\, \frac{a_1}{a_0} = h,\, \frac{a_2}{a_0} = 3,\, \tau = 2\,.$$

Therefore

$$I = \frac{1}{6h}\,[9.2 + (3 + h^2)] = \frac{21 + h^2}{6h}\,.$$

With the given conditions we must find the value of $h = h_{\min}$ for which $I = I_{\min}$. To do this we equate the derivative $\frac{\mathrm{d}I}{\mathrm{d}h}$ to zero:

$$\frac{\mathrm{d}I}{\mathrm{d}h} = \frac{1}{6} - \frac{7}{2h^2} = 0;$$

from this condition we determine the quantity $h_{\min}$:

$$h_{\min} = \sqrt{21}$$

and, therefore,

$$I_{\min} = B_x^* = \frac{\sqrt{21}}{6} + \frac{7}{2\sqrt{21}} = 1.53 .$$

In this case

$$I_{\min\min} = \sqrt{2} \approx 1.41 .$$

Therefore, for $h = \sqrt{21}$ the process is closest to the exponential exponent $-\frac{1}{\sqrt{2}}t$, but is different from it.

(d) *The estimate of how far the process deviates from the extremal process*

In solving practical problems it is often necessary to evaluate the deviation of the real process in the system from that to which it tends, by minimizing the integral estimate.

Let the value of the integral $I_{\min}$ be computed for the selected values of the system parameters.

In addition, let $I_{\min\min}$, the value of this integral estimate for which the transient process coincides with the exponential $x^* = Ce^{-\frac{t}{\tau}}$, be known.

It was shown above that the quantity $I = I_{\min\min}$ is determined by the product of the square of the initial deviation $x^2(0)$ and τ:

$$I_{\min\min} = \tau x^2(0).$$

We consider the difference between these two integral estimates

$$\varepsilon = I_{\min} - I_{\min\min} .$$

Let $\Delta x = x - x^*$, where x is the change of the given coordinate for the chosen values of the parameters, and x^* is the change in this coordinate when $I = I_{\min\min}$.

We put the obtained value of x in the general expression for $I_{\min}$:

$$I_{\min} = \int_0^\infty \left[(x^* + \Delta x)^2 + \tau^2 \left(\frac{dx^*}{dt} + \frac{d\Delta x}{dt}\right)^2\right] dt,$$

or

$$I_{\min} = \int_0^\infty \left\{\left[x^{*2} + \tau^2 \left(\frac{dx^*}{dt}\right)^2\right] + \left[\Delta x^2 + \tau^2 \left(\frac{d\Delta x}{dt}\right)^2\right] + \right.$$

$$\left. + 2\left[x^* \Delta x + \tau^2 \frac{dx^*}{dt} \frac{d\Delta x}{dt}\right]\right\} dt = \int_0^\infty \left[x^{*2} + \tau^2 \left(\frac{dx^*}{dt}\right)^2\right] dt +$$

$$+ \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d\Delta x}{dt}\right)^2\right] dt + 2 \int_0^\infty \left(x^* \Delta x + \tau^2 \frac{dx^*}{dt} \frac{d\Delta x}{dt}\right) dt.$$

In the last integral we put

$$x^* = Ce^{-\frac{t}{\tau}}$$

and, integrating by parts, we can show that the last integral is equal to zero.

The first integral is equal to $I_{\min\min}$. Hence

$$I_{\min} = I_{\min\min} + \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d\Delta x}{dt}\right)^2\right] dt.$$

Taking $I_{\min\min}$ to the left hand side, we obtain

$$\varepsilon = I_{\min} - I_{\min\min} = \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{d\Delta x}{dt}\right)^2\right] dt.$$

Therefore the difference between the two integral estimates is determined by the same integral, except that instead of the variables x and $\frac{dx}{dt}$ it will be necessary to put their increments Δx and $\frac{d\Delta x}{dt}$.

For the subsequent calculation we use the well-known inequality of Bunyakovskii:

$$\int_a^b f_1(t)\, f_2(t)\, \mathrm{d}t \leqslant \sqrt{\int_a^b f_1^2(t)\, \mathrm{d}t \int_a^b f_2^2(t)\, \mathrm{d}t}\,.$$

We use it to estimate the quantity ε. We put $\Delta\, x^2$ in the following form:

$$\Delta x^2 = 2 \int_0^t \Delta x \frac{\mathrm{d}\Delta x}{\mathrm{d}t}\, \mathrm{d}t \leqslant 2 \sqrt{\int_0^t \Delta x^2\, \mathrm{d}t \int_0^t \left(\frac{\mathrm{d}\Delta x}{\mathrm{d}t}\right)^2 \mathrm{d}t}.$$

Multiplying and dividing the right hand side of the equation by τ, we obtain

$$\Delta x^2 \leqslant \frac{2}{\tau} \sqrt{\int_0^t \Delta x^2\, \mathrm{d}t \int_0^t \tau^2 \left(\frac{\mathrm{d}\Delta x}{\mathrm{d}t}\right)^2 \mathrm{d}t}.$$

Since the integrand is positive, the inequality will only be strengthened if we increase the upper limit of integration to infinity:

$$\Delta x^2 \leqslant \frac{2}{\tau} \sqrt{\int_0^\infty \Delta x^2\, \mathrm{d}t \int_0^\infty \tau^2 \left(\frac{\mathrm{d}\Delta x}{\mathrm{d}t}\right)^2 \mathrm{d}t}\,.$$

From the obvious inequality

$$2 \sqrt{ba} \leqslant a + b$$

it follows that

$$\Delta x^2 \leqslant \frac{1}{\tau} \int_0^\infty \left[\Delta x^2 + \tau^2 \left(\frac{\mathrm{d}\Delta x}{\mathrm{d}t}\right)^2\right] \mathrm{d}t = \frac{\varepsilon}{\tau}.$$

Thus, *the deviation of the real transient process from the extremal process does not exceed the quantity:* $\sqrt{\dfrac{\varepsilon}{\tau}}$:

$$|\Delta x| \leqslant \sqrt{\frac{\varepsilon}{\tau}}\,.$$

EXAMPLE 3. In the previous numerical example it was shown that

$$I_{\min\min} = \sqrt{2} \approx 1{\cdot}41 \quad \text{and} \quad I_{\min} \approx 1{\cdot}53.$$

Then

$$|\Delta x| \leqslant \sqrt{\frac{1{\cdot}53 - 1{\cdot}41}{1{\cdot}41}} = \sqrt{\frac{0{\cdot}12}{1{\cdot}41}} \approx 0{\cdot}292.$$

In other words, when the system parameters have been chosen so that the integral estimate

$$I = \int_0^\infty (x^2 + \tau^2 \dot{x}^2)\, dt$$

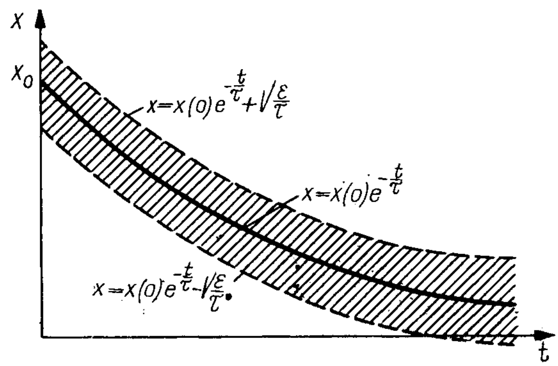


FIG. 170

is minimized, it is ensured that the transient process in the system for these parameters will not go outside the limits of the area lying between the curves (Fig. 170)

$$x = x(0)\, e^{-\frac{t}{\tau}} + \sqrt{\frac{\varepsilon}{\tau}},$$

$$x = x(0)\, e^{-\frac{t}{\tau}} - \sqrt{\frac{\varepsilon}{\tau}}.$$

If any curve lying inside this area, and not only the initial exponential $x = x(0)\, e^{-\frac{t}{\tau}}$, satisfies the technical conditions laid on the transient process, then the selection of the parameters is finished.

The deviation of the resulting process from the exponential process to which it had to approximate by the choice of parameters, proves to be smaller the bigger τ is. Therefore in choosing the exponential to which we want the transient process to approximate, it is not possible to fix τ exceedingly small.

Let it be required by the technical conditions that the transient process does not go outside the limits of the curve shaded in Fig. 171. Of course it can happen that while the initial exponential

$$x = x(0)\, e^{-\frac{t}{\tau}}$$

was chosen so that it does not go outside these limits one of the curves $x = x(0)\, e^{-\frac{t}{\tau}} + \sqrt{\frac{\varepsilon}{\tau}}$ or $x = x(0)\, e^{-\frac{t}{\tau}} - \sqrt{\frac{\varepsilon}{\tau}}$ may do so. Then we must change the value of τ (choose another exponential lying inside the shaded contour) and repeat the calculation.

It is convenient, however, not to fix the value of τ from the very start but to determine all the required quantities in the form of functions of τ. We demonstrate this method by an example.

EXAMPLE 4. We return again to Example 2, but now do not determine h and τ until the end of the calculations, and find them so that they satisfy the given technical conditions in the best way. To do this we express $h_{\min}$, $I_{\min}$ and Δx as functions of τ.

We return to the differential equation

$$\frac{d^2 x}{dt^2} + h \frac{dh}{dt} + 3x = 0$$

with the previous initial conditions:

$$x(0) = 1, \quad \dot{x}(0) = 0.$$

We select the integral estimate

$$I = \int_0^\infty V dt,$$

where $V = x^2 + \tau^2 \dot{x}^2$.

In this case

$$A_1 = 1, \; A_2 = \tau^2, \; x_1(0) = 1, \; x_2(0) = 0.$$

Putting these values in the relations found above, we obtain

$$B_{x_1 x_2} = \frac{1}{3}, \quad B_{x_2} = \frac{\tau^2}{2h} + \frac{1}{6h}, \quad B_{x_1} = \frac{h}{6} + \frac{3\tau^2 + 1}{2h}$$

and, consequently,

$$I = B_{x_1} x_1^2(0) = \frac{h}{6} + \frac{3\tau^2 + 1}{2h}.$$

Taking the derivative $\frac{\mathrm{d}I}{\mathrm{d}h}$ and equating it to zero, we find h as a function of τ^2:

$$h_{\min} = \sqrt{(9\tau^2 + 3)}\ .$$

Putting $h_{\min}$ in the expression for I obtained above, we find:

$$I_{\min} = \sqrt{\tau^2 + \frac{1}{3}}\ .$$

But $I_{\min\,\min} = \tau x_1^2(0) = \tau$, since $x_1(0) = 1$; whence we obtain:

$$|\Delta x| \leqslant \sqrt{\frac{\varepsilon}{\tau}} = \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}\ .$$

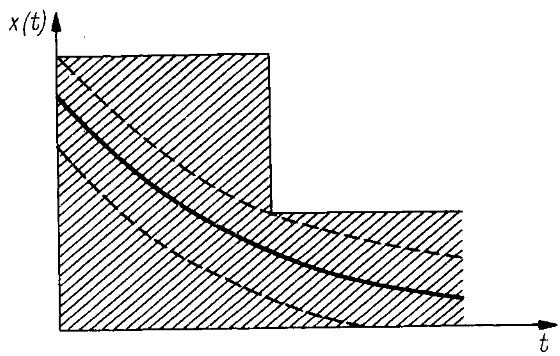


FIG. 171

Now, changing τ, we find $\tau = \tau^*$ for which both curves

$$x = x(0)\,e^{-\frac{t}{\tau}} + \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}$$

and

$$x = x(0)\,e^{-\frac{t}{\tau}} - \sqrt{\sqrt{1 + \frac{1}{3\tau^2}} - 1}$$

stay within the region shaded in Fig. 171.

Then $h = \sqrt{(9\tau^{*2} + 3)}$ is the required value of h.

(e) *The sequence of calculations*

There are two possible sequences of calculations for selecting the system parameters, starting from the integral estimate

$$I = \int_0^\infty (x^2 + \tau^2\,\dot{x}^2)\,\mathrm{d}t.$$

The first sequence is:

1. Choose τ so that the exponential

$$x^* = x(0)\, e^{-\frac{t}{\tau}}$$

satisfies the technical conditions laid on the transient process for the given initial condition $x(0)$ (which is determined by the given disturbance).

2. Find $U(0)$ as a function to the unknown parameters of the system, i.e. those parameters which it is required to select.

3. Minimize this function, i.e. determine the values of the unknown parameters for which the function $U(0)$ attains its absolute minimum $U_{\min}(0)$.

4. Find

$$\varepsilon = U_{\min}(0) - \tau x^2(0)$$

and compute $\sqrt{\frac{\varepsilon}{\tau}}$.

5. Construct the region bounded by the curves $x = x^* + \sqrt{\frac{\varepsilon}{\tau}}$ and $x = x^* - \sqrt{\frac{\varepsilon}{\tau}}$.

If all curves lying in this region satisfy the technical conditions, then the parameters of the system are found. If not, repeat the computation, varying the values of τ. When it proves to be impossible to select τ so that there is any curve in the resulting region which satisfies the technical conditions it is necessary to change other, known, paramters of the system and to repeat the computation.

The second sequence of operations is:

1. Taking τ to be one of the unknown parameters, find $U(0)$ as a function of τ and the other unknown parameters.

2. Minimize this function with respect to the other unknown parameters, i.e. determine for what values of these parameters $U(0)$ attains its minimum, this minimum value of $U(0)$ itself being a function of τ.

3. Find ε as a function of τ.

4. In the x, t-plane, construct the region bounded by the curves

$$x = x(0)\, e^{-\frac{t}{\tau}} + \sqrt{\frac{\varepsilon}{\tau}}$$

and

$$x = x(0)\, e^{-\frac{t}{\tau}} - \sqrt{\frac{\varepsilon}{\tau}}$$

for various values of τ.

If we can find $\tau = \tau^*$ giving a region for which all curves lying in it satisfy the technical conditions laid on the transient process, then the values of the unknown parameters for $\tau = \tau^*$ are taken as the optimal values.

Selecting the parameters on the basis of integral estimates is considerably more worthwhile than selecting them from the degree of stability, although the computations involved in the use of the integral estimates are more tedious.

Despite this, often, particularly in cases when the properties of the system are given by the equations of the process (by the transfer functions), the integral estimates are a much simpler method for the guided selection of the optimal parameters.

7. Estimates of the Process on the Basis of the Form of the Frequency Characteristic

In Section 4 of this chapter it was established that the transient process can be expressed as a function of the real frequency characteristic of the closed system $P(\omega)$ with the help of the integral

$$x(t) = \frac{2}{\pi} \int_0^\infty \frac{P(\omega)}{\omega} \sin \omega t \, d\omega . \tag{4.35}$$

From expression (4.35) it follows that very different systems have only slightly differing transient process when their real characteristics are similar. A more detailed consideration of (4.35) enables us to make some preliminary statements concerning the character and peculiarities of the transient process $x(t)$ on the basis of the form of the real characteristic $P(\omega)$ without having to construct the process itself. Several criteria of this sort are given below. Only some of them can be proved rigorously, the others are obtained from the consideration of a large number of typical examples or deduced for special but very common forms of the characteristics $P(\omega)$. The use of the thirteen criteria

given below therefore requires care, and the results of the investigation must be checked by a construction of the transient process.

We consider the real frequency characteristic of a closed system. Its rough shape is given in Fig. 172. Let

$$P_0 = P(0).$$

A frequency interval over which $P(\omega)$ is positive is called a *positive interval*, and the end-point of this interval will be denoted by ω_P.

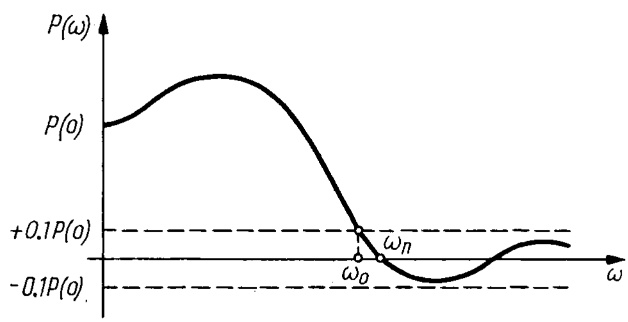


FIG. 172

Parallel to the ω-axis we draw the two straight lines $P = \pm 0{\cdot}1 P_0$ (see Fig. 172). The value of the frequency ω_0, after which the curve $P(\omega)$ does not go outside the strip bounded by these lines, determines roughly the boundary of admissible frequencies of the given system, if it is static. The interval of frequencies bounded by the frequency ω_0 is called the *allowable strip* of the automatic control system. The system will practically not react at all to frequencies lying outside the allowable strip.

Criterion 1. The value of the real frequency characteristic at the point $\omega = 0$ is equal to the value of the coordinate at the end of the transient process, i. e. $P_0 = x(\infty)$.

This property of the real frequency characteristic is easy to prove. From the theory of the Laplace transform it is known that

$$\lim_{p \to 0} pL[x(t)] = \lim_{t \to \infty} x(t) . \tag{4.36}$$

Moreover, for a unit action on the system

$$L[x(t)] = \Phi(p)\, L[\mathbf{1}],$$

or

$$L\,[x\,(t)] = \frac{\Phi\,(p)}{p}\,.$$

Putting this value of $pL\,[x(t)]$ in (4.36) we obtain:

$$\lim_{p\to 0} pL\,[x\,(t)] = \lim_{p\to 0} \Phi\,(p) = \lim_{t\to\infty} x\,(t)\,,$$

or

$$\lim_{p\to 0} \Phi\,(p) = x\,(\infty)\,.$$

But $\mathrm{Im}\,\Phi\,(0) = 0$ and, therefore, $\mathrm{Re}\,\Phi\,(0) = P(0) = x(\infty)$. Therefore, from the above, in an astatic system

$$x(\infty) = P(0) = 0.$$

Criterion 2. In a static system, in order that the size of the overshoot expressed in percentages of the static error does not exceed 18 per cent, i.e.

$$\frac{x_{\max} - x_\infty}{x_\infty} \times 100 < 18\,\%\,,$$

it is sufficient (but not necessary!) that the real frequency characteristic shall be a non-increasing function of the frequency.

We represent the integral (4.35) in the form of a sum

$$x\,(t) = \frac{2}{\pi}\int\limits_0^{\pi/t} \frac{P\,(\omega)}{\omega}\sin\,\omega t\,\mathrm{d}\omega + \frac{2}{\pi}\int\limits_{\pi/t}^{2\pi/t} \frac{P\,(\omega)}{\omega}\sin\omega t\,\mathrm{d}\omega + \dots$$

If the function $P(\omega)$ is always positive and non-increasing, the terms of the series will decrease in magnitude and their sign will be determined by the sign of $\sin\omega\, t$ in the given interval of integration.

From the theory of series it is known that the sum of an alternating decreasing series cannot exceed its first term in magnitude, i. e.

$$x\,(t) \leqslant \frac{2}{\pi}\int\limits_0^{\pi/t} \frac{P\,(\omega)}{\omega}\sin\omega t\,\mathrm{d}\omega\,.$$

The inequality is only strengthened by putting $P(0)$ instead of $P(\omega)$. Then $P(0)$ can be brought out from the integral sign

$$x(t) \leqslant \frac{2}{\pi} P(0) \int_0^{\pi/t} \frac{\sin \omega t}{\omega} d\omega .$$

The last integral is transformed by the substitution

$$\omega t = y, \ d\omega = \frac{dy}{t} .$$

We obtain:

$$x(t) \leqslant \frac{2P(0)}{\pi} \int_0^{\pi} \frac{\sin y}{y} dy .$$

But

$$\int_0^{\pi} \frac{\sin y}{y} dy = 1{\cdot}85 .$$

Therefore

$$x(t) \leqslant \frac{2P(0)}{\pi} \times 1{\cdot}85 = \frac{2}{\pi} \times 1{\cdot}85 x_{\infty} .$$

Subtracting x_{∞} from both sides of the equation we obtain:

$$x(t) - x(\infty) < \left[\frac{2}{\pi} \times 1{\cdot}85 - 1\right] x_{\infty} ,$$

giving

$$\frac{x(t) - x(\infty)}{x(\infty)} \times 100 < \left(\frac{2}{\pi} \times 1{\cdot}85 - 1\right) \times 100 \approx 18\% .$$

In practice the real frequency characteristic usually intersects the w-axis. But by using the given criterion we can often ignore that part of the characteristic which corresponds to frequencies lying outside the allowable strip.

Criterion 3. In order that the transient process shall be monotonic it is sufficient that the real frequency characteristic be a positive function of w with a negative derivative decreasing absolutely in magnitude.

Criterion 4. The process is known to be non-monotonic and to possess overshoot if the condition $|P(\omega)| < P(0)$ is not satisfied for all w

Parallel to the w-axis we draw the two straight lines $P = \pm x(\infty)$. On the basis of this property of the real frequency characteristic we can assert that if the given characteristic goes outside the limits of the strip bounded by the two lines $P = \pm x_\infty$, then the transient process in the system of automatic control will not be monotonic. We note that this criterion is only sufficient, for the real frequency characteristic can remain inside the limits of the indicated region, even when overshoot takes place in the system.

Criterion 5. Overshoot in the system is known to exist, i.e. the process is known to be non-monotonic, if the inequality

$$| P(\omega) | < P(0)\, G(\omega)$$

is not satisfied, where

$$G(\omega) = \cos \frac{\pi}{1 + \dfrac{\omega_p}{\omega}}\,.$$

Criterion 6. If the condition of monotonicity (Criterion 5) is satisfied, then the control time t_0, i.e. the time required for the coordinate x to reach 95 per cent $x(\infty)$, is known to be greater than $\dfrac{4\pi}{\omega_p}$, i. e. to $t_0 > \dfrac{4\pi}{\omega_p}$.

From this criterion it follows that the steeper the curve $P(\omega)$ is, the larger the control time t_0 will be.

Criterion 7. In the general case the control time is known to be greater than $\dfrac{\pi}{\omega_p}$.

Criterion 8. If $P(\omega)$ can be represented as the difference between two non-increasing functions, each of which satisfies Criterion 3, then the overshoot, expressed in percentages, is known to be less than $\left((1{\cdot}18\, \dfrac{P_{\max}}{P_0} - 1\right)$.

i. e. $$\frac{x_{\max} - x(\infty)}{x(\infty)} \cdot 100 < \left(1 \cdot 18\, \frac{P_{\max}}{P(0)} - 1\right) \cdot 100.$$

Criterion 9. Let the transient process $x(t)$ correspond to the real frequency characteristic $P(\omega)$, and the transient process $\overline{x(t)}$ to the real frequency characteristic $S(\omega)$.

Then if $P(n\omega) = S(\omega)$, i.e. if the characteristics differ only in their scale along the w-axis, then the transient processes corresponding to them $x(t)$ and $\overline{x(t)}$ will be connected by the following equation: $x\left(\frac{t}{n}\right) = S(\omega)$.

In other words, if the curves of the real frequency characteristics $P(\omega)$ and $S(\omega)$ only differ in having different scales along the frequency axis, then the transient processes $x(t)$ and $\overline{x(t)}$, corresponding to them, will differ by the same amount in the scale along the time axis. From this criterion follows the criterion: *the more hollow the shape of the real frequency characteristic, the faster will the transient process be concluded.*

To prove this criterion, in (4.35) we put $P(n\omega)$ instead of $P(\omega)$ and find $x(t)$:

$$\overline{x(t)} = \frac{2}{\pi}\int \frac{P(n\omega)}{\omega} \sin \omega t \, d\omega \, .$$

We introduce the new variable

$$\Omega = n\omega$$

Then

$$\overline{x(t)} = \frac{2}{\pi}\int_0^\infty \frac{P(\Omega)}{\Omega/n} \sin \Omega \frac{t}{n} \, d \frac{\Omega}{n}$$

or

$$\overline{x(t)} = \frac{2}{\pi}\int_0^\infty \frac{P(\Omega)}{\Omega} \sin \Omega \frac{t}{n} \, d\Omega .$$

Equating this to (4.35) we obtain at once:

$$\overline{x(t)} = x\left(\frac{t}{n}\right).$$

Criterion 10. If, in the positive interval $0 < \omega < \omega_n$ one real frequency characteristic passes above the other, i.e.

$$P_2(\omega) < P_1(\omega),$$

then the transient processes $x_2(t)$ and $x_1(t)$ corresponding to them will, in the interval of time $0 < t < \frac{\pi}{\omega_n}$, be such that they satisfy the inequality

$$x_2(t) < x_1(t).$$

Criterion 11. If $\frac{P_{max}}{P_0} < 1{\cdot}3$ and $\frac{P(\omega)}{P(0)} \approx 1$ in the frequency interval $\omega < 0{\cdot}1\omega_p$, then the overshoot does not exceed $\sigma = 30$ per cent and the control time is determined by the inequality

$$\frac{\pi}{\omega_p} < T_0 < \frac{8\pi}{\omega_p}.$$

As well as estimating the process on the basis of the real frequency characteristic $P(\omega)$ we sometimes also use the amplitude characteristic $A(\omega) = |W(i\omega)|$.

Criterion 12. If the amplitude characteristic has a high and sharp peak for the frequency $\omega = \overline{\omega}$, then the transient process contains a lightly damped oscillation with frequency $\overline{\omega}$. The steeper and higher the peak the smaller is the damping of this oscillation. The damping is taken to be sufficient if $\frac{A(\overline{\omega})}{A(0)} < \varkappa$. For the values of $\varkappa$ the numbers 1·3, 1·5 and in some cases 2 also, are recommended.

We note that when the characteristic equation of the system has purely imaginary roots $\pm i\overline{\omega}$, i.e. the degree of stability is equal to zero, the amplitude characteristic $A(\omega)$ has, for $\omega = \overline{\omega}$, an infinitely high peak. The estimate of the process made from the height of the peak of $A(\omega)$ is, therefore, closely connected with the estimate of the process made from the degree of stability.

Just as there are no general basic recommendations for the choice of the optimal value of the degree of stability, so there are none for the choice of $\varkappa$. The values given above have been obtained by experiment for systems of concrete types. Criterion 12 can successfully be applied only when the optimal value of $\varkappa$ is known from experience of the design and testing of similar controllers.

Of all these criteria, the greatest value is attached to

Criterion 13. If two systems of automatic control have real frequency characteristics which do not differ significantly, then the transient process in these systems will differ only slightly. The difference in the transient processes will be smaller, the smaller the difference, for low frequencies, between the frequency characteristics.

In fact

$$x(t) = \frac{2}{\pi} \int_0^\infty \frac{P(\omega)}{\omega} \sin \omega t \, d\omega$$

contains under the integral sign the factor $\frac{1}{\omega}$. The difference in $P(\omega)$ will, therefore, affect the magnitude of the integral more for small frequencies than for large. Large differences for large frequencies give small differences in the initial strips of the graphs of the transient processes.

The special feature of this rule consists in the fact that it establishes a criterion for the equivalence of completely different systems from the point of view of the identity of the transient processes taking place in them. Guided by Criterion 13 we can, by investigating simple systems (for example, single-loop systems, consisting of two or three stages) form an index of real characteristics and their corresponding transient processes. Then, when planning a new system, we can select from this index a process which satisfies the given technical conditions and find the typical real frequency characteristic corresponding to it. The problem is then reduced to fitting the real characteristic of the proposed system to this typical characteristic. In many cases the solution of this problem is helped by the use of the other criteria listed in this section and by using logarithmic characteristics.

In those cases when the design office systematically plans controllers of one type, it compiles indexes of its previous designs, using experimental data, and includes in them the characteristics of both those designs which have been completed, and those which have not yet been finished. Such indexes, the concentrated product of the experience of the design office, as they become more complete, become the basic material for the designer, helping him purposefully to plan new controllers.

8. The Analysis of Systems with Random Disturbances Given Statistically

So far we have assumed everywhere in this chapter that the control process is caused by an external disturbance which is a given function of time. Sometimes, not only in special control systems, but also in normal industrial systems, the external disturbance on the system varies continuously and randomly, so that there is no point in analysing the system for a unit disturbance or for any other of the disturbances mentioned above. As an example consider control systems with objects which are subject to continuously changing loads, or the calculation of controllers which are subject to the continuous action of noise, such as the result of the shot effect in the amplifier valves, or of vibrations in chassis, etc. In such cases we take only the statistical characteristics of the external action on the system for our data and the aim of the calculation consists in determining the static characteristics of the process (for example, the deflection in the controlled parameter) which is caused by this disturbance.

The characteristic of a random process

By a *random process* is meant here a random quantity continuously dependent upon time. In particular, a random quantity can be the instantaneous value of the input disturbance on the system, and the output reaction corresponding to it. If we record the random process over a sufficiently large interval of time and compare sufficiently long sections of the record, then they will not coincide, due to the random character of the process. But their statistical characteristics can prove to be identical if the process as a whole does not have a tendency to damping or to intensification. Such a random process is called *stationary*.* In this section the random processes are everywhere assumed to be stationary and we shall not point this out again.

Let $x(t)$ be a random process.

Let us suppose we are given a sufficiently long record of the random process $x(t)$. We divide (Fig. 173) the t-axis into equal intervals of length Δt, and consider the values of $x(t)$ at the points

* The definitions given here of "random process" and "stationary random process" are not exact, and lean more on the intuition of the reader than in exact facts. The exact definition of these concepts will be found by the reader in specialist textbooks.

$t = k\,\Delta t\,(k = 0,\ 1,\ 2,\ \dots,\ N-1)$. The total number of ordinates considered will be equal to N.

Let us fix any value $\varkappa$ and suppose that the number or ordinates at the considered points which are smaller than $\varkappa$ is n. Then n is a function of $\varkappa$; it tends to N for large $\varkappa$ and to zero for small $\varkappa$.

The function

$$F(\varkappa) = \frac{n(\varkappa)}{N}$$

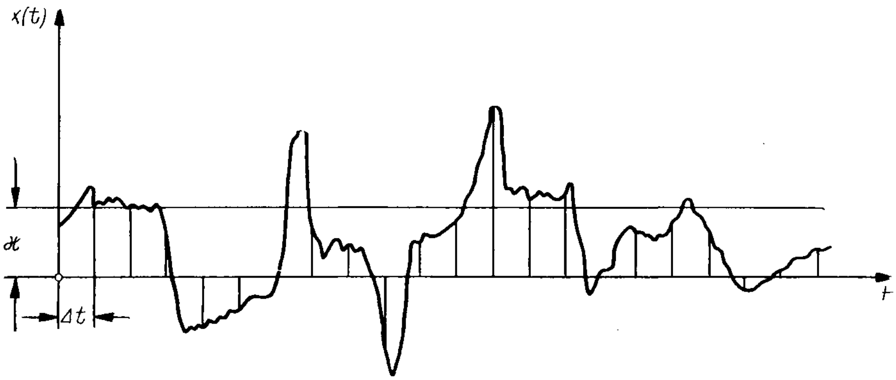


FIG. 173

is called the *distribution function* of the given random process. Its derivative

$$W(\varkappa) = \frac{\mathrm{d}F(\varkappa)}{\mathrm{d}x}$$

is called the *density of the probability distribution*. In Figs. 174 and 175 examples are given of a distribution function and its corresponding distribution density.

We call expressions of the form

$$\bar{x} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)\,\mathrm{d}t \tag{4.37}$$

and

$$\bar{x}^2 = \frac{1}{2T} \int_{-T}^{T+} x^2(t)\,\mathrm{d}t \tag{4.38}$$

respectively the *mean* and the *mean square* of the process $x(t)$.

Let us now consider a function of the real variable τ, defined by the equation

$$R_x(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)\, x(t)\, \mathrm{d}t. \tag{4.39}$$

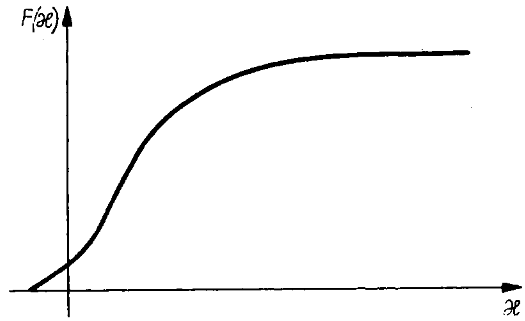


FIG. 174

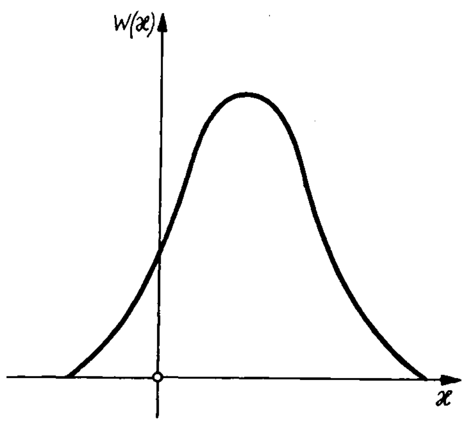


FIG. 175

$R_x(\tau)$ is called the *auto-correlation function*. In its place we often consider the *normalized auto-correlation function* defining it by the equation

$$r_x(\tau) = \frac{R_x(\tau)}{R_x(0)}. \tag{4.40}$$

If two random processes $x(t)$ and $y(t)$ are given, then the functions

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)\, y(t+\tau)\, dt \tag{4.41}$$

and

$$r_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sqrt{R_x(0)\, R_y(0)}} \tag{4.42}$$

are called respectively the *mutual correlation function* and the *normalized mutual correlation function.*

If a sufficiently long record of the random process is given, the auto-correlation function can be calculated immediately from formula (4.39), by replacing the integral by a finite sum and successively giving different values to τ. Special machines called correlators can be used for this purpose. The recorded process is input into the machine on paper or magnetic tape, and the machine calculates immediately the graph of the function or its values for different values of τ.

For this purpose it is convenient to copy the oscillogram of the process on to tracing-paper and then lay the tracing-paper on the oscillogram, displacing it by the amount τ along the t-axis. The calculation of the correlation function then reduces to the multiplication of corresponding ordinates and the summation of the products.

In the analysis of automatic control systems the following properties of autocorrelation functions are important:

1. *The value of the auto-correlation function for* $\tau = 0$ *is equal to the mean square* $\overline{x}^2$

$$R_x(0) = \bar{x}^2. \tag{4.43}$$

This follows immediately from the definition of $R_x(\tau)$ and $\overline{x}^2$.

2. *The value of the auto-correlation function for* $\tau \neq 0$ *does not exceed its value at* $\tau = 0$, *i.e.*

$$R_x(\tau) \leqslant R_x(0) \quad \text{or} \quad r_x(\tau) \leqslant 1. \tag{4.44}$$

To prove this property let us consider the obvious inequality

$$[x(t) \pm x(t+\tau)]^2 \geqslant 0$$

or

$$x^2(t) + x^2(t+\tau) \geqslant 2x(t)\, x(t+\tau).$$

We integrate the right- and left-hand sides with respect to t from $-T$ to $+T$, divide these integrals by $2T$ and pass to the limit as $T \to \infty$:

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x^2(t)\,dt + \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x^2(t+\tau)\,dt \geqslant$$
$$\geqslant 2 \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)\,x(t+\tau)\,dt.$$

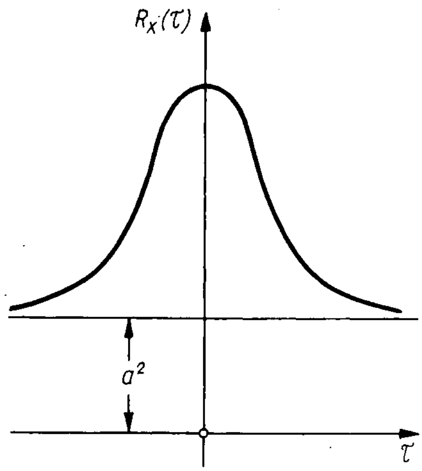


FIG. 176

We may take $t+\tau$ in the second term on the left hand side of this inequality as a new time t^*; then $dt = dt^*$ and we see at once that the second term is equal to the first, which is equal to the mean square. On the right-hand side of the inequality we have twice the auto-correlation function. Therefore

$$2\,\bar{x}^2 \geqslant 2R_x(\tau)$$

and on the strength of the first property $R_x(0) \geqslant R(\tau)$.

3. *If the given process does not contain non-random constituents, then $R_x(\tau) \to 0$ as $\tau \to \pm\infty$. But if $x(t)$ is the sum of a random process and a constant constituent, a, $R_x(\tau) \to a^2$ as $\tau \to \pm\infty$ (Fig. 176). If, finally, $x(t)$ is the sum of a random and a given periodic process, in the*

limit as $\tau \to \pm\infty$ the function $R_x(\tau)$ becomes a periodic function with the same frequency (Fig. 177).

This property follows immediately from the definition of the auto-correlation function and from the fact that the auto-correlation function for $x(t) \equiv a$ is a^2, and for $x(t) = \sum_{k=1}^{n} a_k \sin(k\omega t + \varphi_k)$ is $\sum_{k=1}^{n} a_k^2 \cos k\omega$.

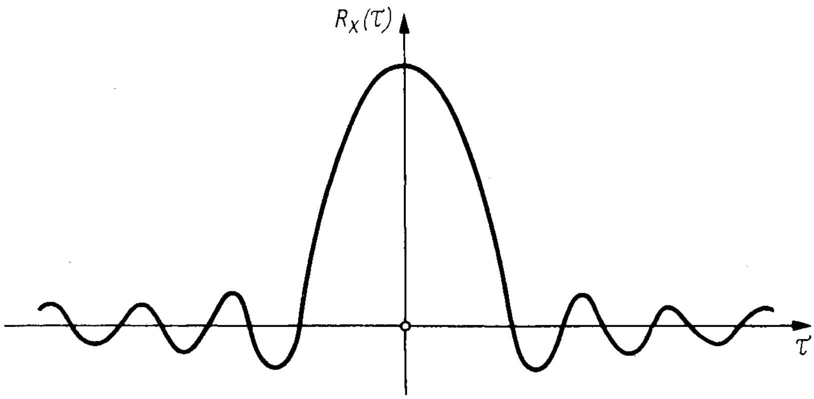


FIG. 177

Processes for which a random action is added to a constant or a periodic action are encountered very frequently and it is difficult to establish the presence of both actions from the record of the process. The construction of the correlation functions enables us at once to isolate these cases and by removing the non-random constituent, to obtain the correlation function for the random process alone.

We now return to a consideration of the random process $x(t)$ itself.

The function $x(t)$ has no Fourier transform, since as $t \to \infty$ it does not tend to zero. We therefore introduce the function $x_T(t)$ which is equal to $x(t)$ for $T \leqslant t \leqslant T$ and is equal to zero for all other values of t (Fig. 178). The Fourier transform of this function exists; it is bounded and tends to zero as $t \to \infty$:

$$\Phi_{x_T}(i\omega) = \int_{-T}^{+T} x_T(t)\, e^{-i\omega t}\, \mathrm{d}t. \tag{4.45}$$

Then we can consider the function

$$\Phi_x^*(i\omega) = \lim_{T\to\infty} \frac{1}{2T} \Phi_{x_T}(i\omega) \tag{4.46}$$

as the *conditional Fourier transform* of the random process $x(t)$.

The square of the modulus of this complex function

$$S(\omega) = |\Phi_x^*(i\omega)|^2 \tag{4.47}$$

is called the *spectral density of the random process* $x(t)$.

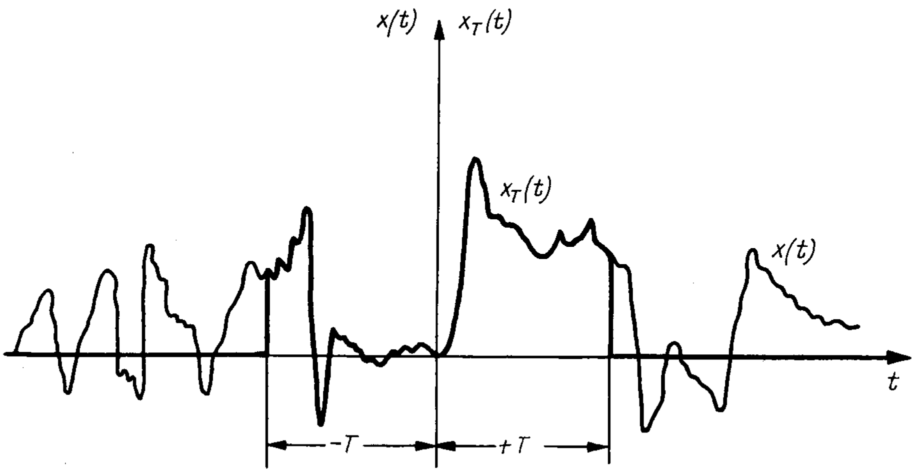


FIG. 178

By definition the correlation function of $x_T(t)$ is of the form

$$R_T(\tau) = \frac{1}{2T} \int_{-T}^{T} x_T(t)\, x_T(t+\tau)\, \mathrm{d}t \quad \text{for} \quad -T < t < T, \tag{4.48}$$

$R_T = 0$ for $t < -T$ and $t > +T$, while the correlation function of $x(t)$ is equal to

$$R(\tau) = \lim_{T\to\infty} R_T(\tau). \tag{4.49}$$

The function $R_T(\tau)$ has the Fourier transform

$$\Phi_{R_T}(i\omega) = \int_{-\infty}^{\infty} R_T(\tau)\, e^{-i\omega\tau}\, \mathrm{d}\tau = \frac{1}{2T} \int_{-\infty}^{\infty} \left[\int_{-T}^{+T} x_T(t)\, x_T(t+\tau)\, \mathrm{d}t \right] e^{-i\omega\tau}\, \mathrm{d}\tau.$$

In the integral between square brackets, the limits of integration can be widened to $-\infty$ and $+\infty$ without changing the result of the integration, since the integrand is equal to zero for $t < -T$ and $t > T$.

Then

$$\Phi_{R_T}(i\omega) = \frac{1}{2T}\int\limits_{-\infty}^{\infty} \mathrm{d}\tau \int\limits_{-\infty}^{\infty} x_T(t)\, x_T(t+\tau)\, e^{-i\omega(t+\tau)}\, e^{i\omega t}\, \mathrm{d}t =$$

$$= \frac{1}{2T}\int\limits_{-\infty}^{\infty} x_T(t)\, e^{i\omega t}\, \mathrm{d}t \int\limits_{-\infty}^{\infty} x_T(t+\tau)\, e^{-i\omega(t+\tau)}\, \mathrm{d}\tau.$$

Introducing the new variable $t + \tau = \lambda$ and putting $t = \text{const}$ in the second integration (with respect to τ), we obtain:

$$\Phi_{R_T}(i\omega) = \frac{1}{2T}\int\limits_{-\infty}^{\infty} x_T(t)\, e^{i\omega t}\, \mathrm{d}t \int\limits_{-\infty}^{\infty} x_T(\lambda)\, e^{-i\omega\lambda}\, \mathrm{d}\lambda.$$

Recalling the relation (4.45), we find that the function $\Phi_{R_T}(i\omega)$ is real, and not complex, and is equal to

$$\Phi_{R_T}(\omega) = \frac{1}{2T}\, \Phi_{x_T}(i\omega)\, \Phi_{x_T}(-i\omega) = \frac{1}{2T}\, |\Phi_{x_T}(i\omega)|^2\,. \qquad (4.50)$$

We now take the limit of the left- and right-hand sides as $T \to \infty$:

$$\lim_{T\to\infty} \Phi_{R_T}(\omega) = \lim_{T\to\infty} \frac{1}{2T}\, |\Phi_{x_T}(i\omega)|^2\,. \qquad (4.51)$$

On the strength of (4.46) and (4.47) the right hand side is equal to $S(\omega)$. We can consider the left-hand side as the conditional Fourier transform* of the correlation function $R_T(\tau)$, just as the conditional Fourier transform of the random process $x(t)$ was introduced above.

* If the given process $x(t)$ is such that $R(\tau)$ tends to zero as $\tau \to \infty$, then there is no need to introduce the conditional Fourier transform for $R(\tau)$ and $\Phi_R^*(\omega) = \Phi_R(\omega)$ is the Fourier transform in the usual sense. Bearing this case in mind, we shall speak of the Fourier transform of the function $R(\tau)$ and omit the asterisk in $\Phi_R(\omega)$.

Then

$$\Phi_R(\omega) = \int_{-\infty}^{\infty} R(\tau)\, e^{-i\omega\tau}\, d\tau = S(\omega) \tag{4.52}$$

and, conversely,

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, e^{+i\omega\tau}\, d\omega. \tag{4.53}$$

Remembering that $R(\tau)$ is an even function, and that $S(\omega)$ is real, we can omit the imaginary parts which equal zero, and replace

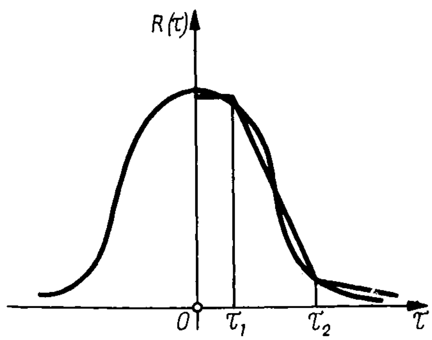


FIG. 179

the integral from $-\infty$ to $+\infty$ by twice the integrals from 0 to $+\infty$.
Then we obtain

$$\boxed{S(\omega) = 2 \int_{0}^{\infty} R(\tau) \cos \tau\omega \, d\tau} \tag{4.54}$$

and

$$\boxed{R(\tau) = \frac{1}{\pi} \int_{0}^{\infty} S(\omega) \cos \tau\omega \, d\omega.} \tag{4.55}$$

i.e. *the correlation function and the spectral density are the cosine Fourier transforms of one another.*

Formula (4.54) is used to construct the spectral density directly from the given correlation function. To do this it is necessary to replace the curve $R(\tau)$ by a broken straight line or segments of an ex-

ponential, after which it is easy to carry out the integration (4.54) for the sections $0 - \tau_1$, $\tau_1 - \tau_2$, ..., (Fig. 179).

Often, in concrete problems, we can approximate to the function $R(\tau)$ for $0 < \tau < \infty$ by the sum of several simple functions, such as exponentials, and, by putting them in (4.54), compute the integral.

Now putting $\tau = 0$ in (4.55) and remembering (4.43) we obtain:

$$\boxed{\bar{x}^2 = \frac{1}{\pi}\int\limits_0^\infty S(\omega)\,d\omega.} \tag{4.56}$$

i.e. the mean square of the random quantity is equal to the integral of spectral density, along the semi-axis $\omega > 0$, divided by π.

The connexion between the spectral density of random processes at the input and the output of a linear system

Let an external action act at the input of a linear system, and let this action be a stationary random process. As a result of this disturbance the output coordinate of the system will change with time and the function $x_{\text{out}}(t)$ will then also be a stationary random process. It is required to find the statistical characteristic of the process at the output of the system, knowing the statistical characteristic of the random disturbance.

The most simple relation is that established between the spectral density of the disturbance $S_1(\omega)$ and the spectral density of the process at the output of the system $S_2(\omega)$.

In order to find it, put

$$x_1 = x_{\text{out}} \text{ and } x_{1T} = x_{T\text{out}} = \begin{cases} x_{\text{out}} & \text{for } -T < t < T, \\ 0 & \text{for } t < -T \text{ and } t > T, \end{cases}$$

$$x_2 = x_{\text{in}} \text{ and } x_{2T} = x_{T\text{in}} = \begin{cases} x_{\text{in}} & \text{for } -T < t < T, \\ 0 & \text{for } t < -T \text{ and } t > T. \end{cases}$$

Then there exist Fourier transforms for x_{1T} and x_{2T} and these are related by the general formulae of linear control theory

$$\Phi_{x_{1T}} = \Phi_{x_{2T}} \cdot \Phi(i\omega),$$

where $\Phi(i\omega)$ is the amplitude-phase characteristic of the closed system from x_2 to x_1.

Dividing this equation by $2T$ and taking the limit as $T \to \infty$ on the strength of (4.46) we find:

$$\Phi^*_{x1T} = \Phi^*_{x2T}\,\Phi(i\omega),$$

whence

$$|\Phi^*_{x1T}|^2 = |\Phi^*_{x2T}|^2\,|\Phi(i\omega)|^2.$$

Recalling the definition of the spectral density (4.47) we obtain:

$$\boxed{S_{x_1}(\omega) = S_{x_2}(\omega)\,|\Phi(i\omega)|^2,} \tag{4.57}$$

i.e. *the spectral density of a random process at the output of the linear system is equal to the spectral density of the external disturbance at the input, multiplied by the square of the modulus of the frequency characteristic from the input to the output of this system.*

The sequence of calculations for a system when the disturbance is a stationary random process

We begin the calculation by processing a sufficiently long record of the disturbance. With the aid of a correlator or by direct calculation (in this case the integral being replaced by a finite sum) we find from the record the correlation function of the disturbance. If this correlation function does not tend to zero as $\tau \to \infty$, but tends to the value $R(\tau) = a^2$ then the mean of the disturbance is equal to a. In this case it is necessary to shift the origin in the calculation of the disturbances to a, and in the correlation functions to a^2, so that in the new coordinates the mean of the disturbance and the value of the correlation function will equal to zero as $\tau \to \infty$.

From the correlation function found in this way, we construct the curve of the spectral density. To do this we replace the correlation function by a broken line (or we use some other approximation) and then compute the integral in (4.54) by parts.

Then, using the basic relation (4.57) we find the spectral density of the random process at the output of the system, first having constructed the graph of the square of the modulus of the amplitude-

phase characteristic as a function of the frequency, and multiply the ordinates of this graph by the values of the spectral density of the disturbance at the same frequencies.

Using a planimeter on the constructed curve of the spectral density of the output from formula (4.56) we can find the mean square deviation of the coordinate of the system for the given input random process.

The greatest time in this calculation is spent in determining the spectral density of the input disturbance.

The calculations needed in the determination of the spectral density of the input signal can, however, be mechanized by the use of correlators and simplified by satisfactory representation of the correlation function in the form of a sum of several elementary functions (lines, exponentials, etc.).

9. Concluding Remarks

It is, of course, not possible to construct the control process or to estimate its course by linear methods if, due to the technical conditions, it is necessary to carry out the calculation for large disturbances (for example for a total break in the load on the object, or for a sharp displacement of the tuning element from one extreme position to the other, and so on). The use of linear methods in the construction and evaluation of the process is only justified in three cases:

(a) When the control process is described by linear equations not only for small disturbances, but also for any given disturbance (in this case the use of linear methods enables us to design the system completely).

(b) When only some of the technical conditions must be satisfied for sufficiently small disturbances (in this case the linear analysis is of a limited value and is used to ensure part of the technical conditions).

(c) When the technical conditions demand that the process satisfies the given requirements for any disturbances, including small ones (in such cases it is necessary first of all, using linear methods, to ensure the required course of the process for sufficiently small disturbances, and only then to verify that the requirements on the process are also satisfied for large disturbances).

The sequence of the calculations when using linear methods is determined mainly by the form in which the initial material is given:

whether in the form of the equations of motion (the transfer function) or in the form of frequency characteristics.

If we start from the equations of motion, then it is most convenient to use integral estimates for the preliminary choice of the parameters.

After using these indirect estimates to select optimal values of the parameters we calculate the roots of the characteristic equation, and from the Heaviside formula calculate the transient process. If it does not satisfy the technical conditions, then it is necessary to change the values of the parameters in some way and once again to calculate the roots of the characteristic equation, and from them the process. The work involved in constructing the process for fixed values of the parameters can often be made easier by the use of the graphical methods described in Section 3 of this chapter. It is especially convenient and simple for single-loop circuits not containing a large number of oscillatory stages and including other stages with similar time constants.

When the initial material for the calculation is given in the form of the frequency response characteristic the calculation begins with the preliminary estimate of the process from the real and amplitude frequency characteristics. Then, two methods are possible. If sufficient information has been gained from the frequency characteristics of similar systems and from their corresponding transient processes, we must choose a frequency characteristic which meets the requirements of the transient process and which satisfies the technical data. After this the system parameters and the stabilizing means are chosen so that the frequency characteristic of the given system approximates to the selected one. Checking is carried out by constructing the process from the resulting real characteristic (by dismembering it into trapezia and using the tables of the h-function).

When sufficient material about the frequency characteristics of similar systems is not available, we can now only choose the initial parameters for the evaluation of the process by indirect methods. Then, by dividing the characteristic up into trapezia and using the h-function tables, we construct the process. If it does not satisfy the technical conditions, then the parameters must be changed (using the indirect criteria given in Section 7), and the real frequency characteristic and the process must once again be constructed.

The total coefficient of amplification of the system K is almost

always one of the selected parameters. In these cases it is convenient to reconstruct the frequency characteristic from the D-partition curve for K.

Thus, if there is not much experimental material for similar designs and systems, then both in the calculation which makes use of the equations of the process, and in the calculation which uses the frequency characteristics, we must find the optimal parameters by selection, each time testing the choice of the parameters by constructing the process. Indirect estimates of the process which guide us in selecting the initial parameters make it much easier.

Various methods have been developed for the synthesis of a linear control system for the given technical conditions and for the synthesis of stabilizing assemblies. These very attractive methods still require a great deal of testing in their practical computational operations and are not described in this book, since in the synthesis of a system for given technical conditions non-linearities have to be taken into account even more than in its analysis.

The calculation of the control process becomes considerably more complicated in those very common instances when the technical data lays bounds on the process for large disturbances, or when the equations for large disturbances are non-linear, or even more in those cases when the system contains non-linearizable elements and their presence cannot be disregarded. This relates, in particular, to general industrial controllers: pneumatic universal controllers of almost all types which are now very widespread. They usually contain two non-symmetric relays (such as in Fig. 27 and 28). When the processes taking place are controlled slowly, i.e. when the time taken to set up the object of control is incomparably greater than the setting-up time in the controller, we can ignore all the time constants and damping in the controller and consider it as if it were a device instantaneously realizing the required laws of the controlling action.

If this law is linear, then we can use linear methods. But very often (for example, in the control of outflow, pressure, etc.) the times needed to set up the object are of the same order as those needed for the controller. Linear methods are then not applicable, and to analyse the system we must use non-linear methods, which have been studied considerably less.

In the simplest cases, when the non-linear elements can be isolated, the process can be constructed by the graphical method de-

scribed in Section 3. The parameters of the system must then be chosen at random, by trial an error, each time checked by constructing the process. In the more complicated cases various other methods of numerical or graphical integration of non-linear differential equations are used. These methods are described in many courses on numerical and graphical methods of approximative analysis, and therefore usually their study is not included in courses on automatic control theory. Up to now we have not succeeded in making use of the specific character of the differential equations describing the processes in automatic control systems to simplify these general methods very much or to determine any other indirect estimates of the control process in non-linear systems from them which are suitable for technical calculations.

CHAPTER V

AUTO- AND FORCED OSCILLATION IN NON-LINEAR SYSTEMS

1. General Remarks Concerning Periodic States in Non-Linear Systems

In Chapter II equations describing the control process in a real system (the initial equations) were derived, but they were then simplified and replaced by linear differential equations with constant coefficients. In this way the analysis of the real system was replaced by the analysis of its linear model.

In our consideration of the conditions for the stability of the linear model it was established that its stability or instability depended only on the properties of the system and was completely independent of the magnitude of the initial deviation, and that in an unstable system the deviations of all the generalized coordinates grew without limit whatever the initial conditions.

If we do not consider systems whose parameters correspond exactly to the boundary of the region of stability, then in the linear model only two types of motion turn out to be possible. In a stable linear model, after any initial deviation, the values of the generalized coordinates tend over the course of time (monotonically or otherwise) to the values corresponding to the position of equilibrium. In an unstable linear model, on the contrary, after any initial deviation, the values of the generalized coordinates monotonically or non-monotonically increase in absolute magnitude. But when the parameters correspond exactly to the boundary of the region of stability, undamped oscillations are possible. The amplitude of these oscillations will depend on the initial conditions. A very slight change in the parameters will cause the oscillations either to become damped or to increase without limit. No other motions are possible in the linear model.

Observation of real automatic control systems has shown that the motions in them are considerably more varied than those in the linear model.

In real systems undamped oscillations will often occur. These oscillations possess a definite stability: after the disturbance they are restored after a certain time, i.e. both the form of the oscillations and their frequency are restored. We can change the form and frequency of these oscillations by changing the system parameters.

It is not only with regard to the possibility of undamped oscillations that the motion in a real system differs from that in its linear model, for in real systems, as opposed to linear models, the character of the motion often depends on the magnitude of the initial deviation. In real systems there can exist a threshold for the initial deviations such that initial deviations staying within it cause motions which coincide with the position of equilibrium, while after initial deviations which go above the given threshold, stable undamped oscillations are set up in the system.

In a number of automatic control systems not only one, but several states of undamped oscillations are possible, where it depends only on the magnitude of the initial deviations which of these oscillations is set up in the system. Thus it can frequently be observed, for example, that after small initial deviations in the system high-frequency undamped oscillations are set up, which are restored after an initial deviation which does not exceed the defined threshold. But if the initial deviations do exceed this threshold, low-frequency undamped oscillations of considerably larger amplitude are set up in the system.

A phenomenon of this sort can be caused only by factors which are not taken into account in the consideration of the linear model. These factors are the non-linearities which were replaced by linear relations (in the case of linearizable non-linearities) or else were completely ignored (in the case of non-linearizable non-linearities) on transition to the linear model. To describe the above motion and, in particular, undamped oscillations, by means of equations we have now to take the non-linearities into account.

In the absence of external periodic actions the undamped oscillations in the automatic control systems about which we have spoken arise only because of the internal properties of the control system. Their frequency as a whole is determined by the properties of the

system and changes when its parameters change. These are the characteristic *auto-oscillations*, which arise as a result of the equality between the losses in energy during the oscillatory cycle and the influx of energy from outside, non-oscillatory, sources. Usually the controlled object or amplifiers provide this external source of energy. It is only because of the presence of non-linearities that this balance of energy in the oscillatory cycle is possible, and the calculation of the conditions for the existence of undamped oscillations reduces in essence to the determination of the conditions for the realization of this balance.

The basic content of this chapter consists of methods serving to determine the parameters of the auto-oscillations and the conditions for their existence. For this we shall consider the question of the search for periodic solutions of the initial non-linear equations describing the control process.

Periodic solutions are important not only because some of them (namely, the stable periodic solutions) correspond to auto-oscillatory states. In some of the simplest cases a knowledge of the periodic solutions enables us to estimate all the varied motions possible in the given system, and the conditions for existence of periodic solutions enable us to determine the values of the parameters for which the character of the motion changes.

We first consider only systems of equations which differ from linear systems by the presence in one of the equations (the first, say) of a non-linear function of one of the coordinates.

A system of this kind can often be reduced to the form*

$$\dot{x}_1 = \sum_{j=1}^{n} a_{1j} x_j + f(x_k), \quad \dot{x}_i = \sum_{j=1}^{n} a_{ij} x_j, \tag{5.1}$$

where $i = 2, \ldots, n$ and k is any number from 1 to n.

Putting $y = f(x_k)$ and assuming at present that $f(x_k)$ is a continuous, smooth function (i.e. with a continuous derivative), we eliminate all the coordinates apart from y and x_k (and can now omit the

* In control theory we also come across cases wher the system of equations (5.1) contains several non-linear functions or non-linear functions of several variables. Such cases will be considered very briefly below.

index k). As a result we obtain

$$\left.\begin{aligned} a_0\frac{d^n x}{dt^n}+a_1\frac{d^{n-1}x}{dt^{n-1}}+\ldots+a_n x &= \\ = b_0\frac{d^m y}{dt^m}+b_1\frac{d^{m-1}y}{dt^{m-1}}&+\ldots+b_m y, \\ y = f(x), & \end{aligned}\right\} \quad (5.2)$$

where a_j and b_j are constant coefficients which are obtained by the usual procedure for the elimination of superfluous coordinates*.

Starting from expressions which we shall explain below, in Section 1—7, it will everywhere be assumed that the open system is stable, i.e. that all the roots of the equation

$$a_0 p^n + a_1 p^{n-1} + \ldots + a_n = 0$$

lie to the left of the imaginary axis.

The given system can now be represented by a conditional linear stage with transfer function $W(p) = \frac{K(p)}{D(p)}$ where

$$D(p) = a_0 p^n + a_1 p^{n-1} + \ldots + a_n,$$
$$K(p) = b_0 p^m + b_1 p^{m-1} + \ldots + b^m,$$

which is closed by non-linear feedback (Fig. 180). It is required to determine the periodic solutions of this system of equations.

Up to the present time, methods for the exact solution of this problem have not been found, even for the simplest special case, if the degrees of $D(p)$ and $K(p)$ are not restricted and if $f(x)$ is an arbitrary function (or even a continuous function).

If no restrictions are placed on the degrees of $D(p)$ and $K(p)$, the problem of determining the periodic states can only be solved exactly for a few cases when the characteristic $f(x)$ consists only of straight line segments. Relay systems and some systems which are non-linear because of the forces of dry friction are typical systems of this kind.

* For more detail see Section 3 and Appendix 1.

But even when $f(x)$ is a piecewise-linear function, the problem of determining the periodic states reduces to the solution of a system of transcendental equations; it is possible to indicate a general method for the solution of these equations only in the very special case when $f(x)$ is a symmetric relay Z-characteristic and when the most simple symmetric periodic state is considered.

In no other cases have exact methods for the determination of periodic solutions been found and we must have recourse to a variety of approximate methods.

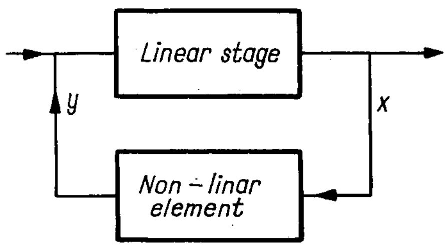


FIG. 180

For this reason the remaining contents of this chapter can be divided into two large parts. In the first part we explain approximate methods of determining the periodic states in cases when they are nearly harmonic, and in the second part we explain the exact determination of periodic states in systems with a piecewise linear characteristic for the non-linear element.

A. THE APPROXIMATE DETERMINATION OF PERIODIC STATES WHICH ARE NEARLY HARMONIC

2. The Conditions for Which Periodic States are Nearly Harmonic

In Section 3 we shall consider one of the methods of approximate determination of periodic solutions based on the method of harmonic balance.

If we are only using it to determine periodic solutions, this method is based on the assumption that the periodic solution is nearly har-

monic and that we can therefore look for it in the form

$$x = A \sin \omega t \,,$$

where A and ω are the amplitude and frequency of the required solution.*

To discover under what conditions this assumption is legitimate, and so to establish the limits within which this approximate method can be validly applied to problems of automatic control, we will dwell on the question of the propagation of a harmonic oscillation in the system represented in Fig. 180.

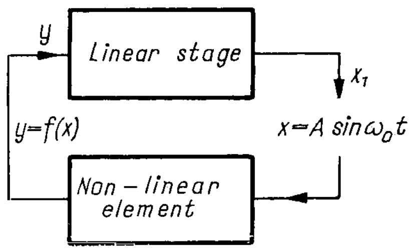


FIG. 181

We open the system (Fig. 181) and suppose that the harmonic action

$$x = A \sin \omega_0 t$$

is applied at the input of the system. Then the coordinate y changes according to the law

$$y = f(A \sin \omega_0 t) \,.$$

The function $f(A \sin \omega_0 t)$ is periodic. We expand it in a Fourier series

$$y = f(A \sin \omega_0 t) = y_0 + \sum_{k=1}^{n} A_k \sin (k\omega_0 t + \varphi_k)$$

and represent it by a line spectrum, drawing lines proportional to the amplitudes corresponding to the harmonics y_0, A_1, A_2, ... at the points 0, ω_0, $2\omega_0$, $3\omega_0$, ... (Fig. 182).

* The phase is assumed to be zero. This does not limit the generality of the solution, since when the equation of motion does not contain t it is clear that this can always be made true by the suitable choice of the time origin.

Of course, harmonics (and their corresponding lines in the spectrum) need not exist for all frequencies which are multiples of ω_0. For instance, if the characteristic $f(x)$ is odd, then the amplitudes of all the even harmonics (0, $2\omega_0$, $4\omega_0$ etc.) are equal to zero.

Consequently, in the linear part of the system there acts an oscillation whose spectrum is represented in Fig. 182 for the general case when all the harmonics are present.

On the strength of the principle of superposition, the action of each harmonic is independent of the others. In accordance with the

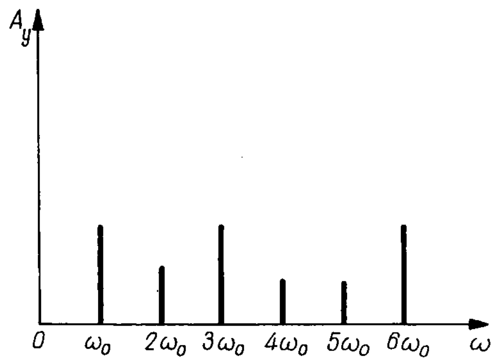


FIG. 182

general laws concerning the propagation of an oscillation through a linear system, the oscillation

$$x_1 = y_0 \frac{K(0)}{D(0)} + \sum_{k=1}^{n} A_k^* \sin(k\omega_0 t + \varphi_k^*), \tag{5.3}$$

is set up at the output of the linear part of the open system, where

$$A_k^* = A_k \left| \frac{K(ik\omega_0)}{D(ik\omega_0)} \right|,$$

$$\varphi_k^* = \varphi_k + \arg \frac{K(ik\omega_0)}{D(ik\omega_0)} .$$

Thus, when a harmonic oscillation is applied at the input of an open system, non-harmonic oscillations are set up at the output.

The spectrum of these oscillations (Fig. 183) will contain the same frequencies as the spectrum of the oscillations of the coordinate y, but the length of each line (the amplitude of the harmonic) will be changed by a factor of $\left|\frac{K(ik\omega_0)}{D(ik\omega_0)}\right|$.

The amplitude characteristic of the linear part of the system

$$A = \left|\frac{K(i\omega)}{D(i\omega)}\right|$$

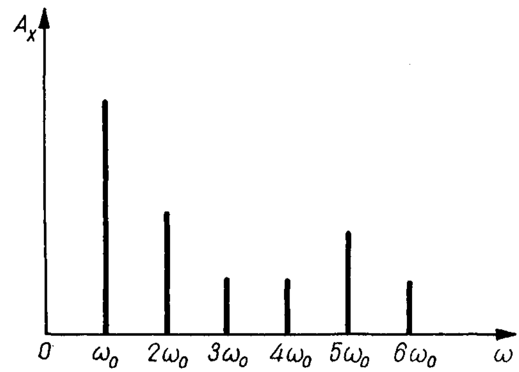


FIG. 183

enables us to establish at once how much the amplitudes of all the harmonics caused by non-linearities change when they are passed through the linear part of the system. To do this it is necessary, for a given value of the basic harmonic ω_0, to consider the ordinate of the amplitude characteristic at the points $\omega = \omega_0$, $\omega = 2\omega_0$, $\omega = 3\omega_0$ and so on.

From the above reasoning it follows that the *oscillations of the coordinate x_1 (and even more, of y) which can be set up in a closed system are always non-harmonic.*

We can, however, find conditions for which the oscillations are nearly harmonic. There are two groups of conditions of this kind, and we shall consider them separately.

1. We assume (although this cannot actually happen) that the linear part of the system is a filter, completely blocking the frequencies $\omega > \omega_c$, i. e. that its amplitude characteristic for $\omega > \omega_c$ coincides with the ω-axis (Fig. 184). Suppose, further, it is known that the

non-linear element gives rise to harmonics beginning with the Sth (for example, for an odd characteristic $S = 3$, or for an arbitrary characteristic $S = 2$). Then, whatever the remaining part of the characteristic $y = f(x)$ does, the oscillations in the system will be strictly harmonic provided their frequency ω_0 satisfies the condition

$$\frac{\omega_c}{S} < \omega_0 < \omega_c.$$

In fact, with this condition all the harmonics produced by the non-linear element are blocked by the linear part of the system.

If the amplitude characteristic were as in Fig. 184, we could look for periodic solutions of the form

$$x = A \sin \omega t \tag{5.4}$$

and, finding the frequency $\omega = \omega_0$, verify whether it satisfied the inequality

$$\frac{\omega_c}{S} < \omega_0 < \omega_c.$$

If the resulting frequency satisfies this inequality, then, in addition, the initial assumption (5.4) is valid. If not, then this assumption as false, and it is not possible to find a periodic solution in the form (5.4).

The linear part of a real system cannot have an amplitude characteristic like the one in Fig. 184, since $A > 0$ for all frequencies.

But, remembering what we said in Section 7 of Chapter IV about the frequency of cut-off, we can usually ignore that part of the characteristic which lies to the right of the cut-off frequency, i.e. the real characteristic can come as near as we please to the characteristic of an exact filter which blocks all high frequencies with $\omega > \omega_c$ (Fig. 185).

In these cases, just as in the case of an exact filter, it is natural to assume that the oscillations are nearly harmonic, if

$$\frac{\omega_c}{S} < \omega_0 < \omega_c.$$

We call this assumption the *filter hypothesis*.

Of course, the conditions of a filter can only be realized if the linear part of the system is stable. If a non stable stage (not embraced by feedback) is present in the linear part of the system, when a harmonic disturbance is applied at the input of this part, no harmonic oscillation is set up at its output, since the constituents of the free oscillations are superimposed. Therefore the idea of a "filter" in an unstable network has no physical meaning, and so there is no meaning and no point in substantiating that the oscillations are nearly harmonic on account of the linear filter.

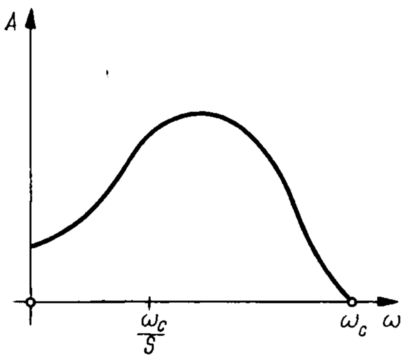


FIG. 184

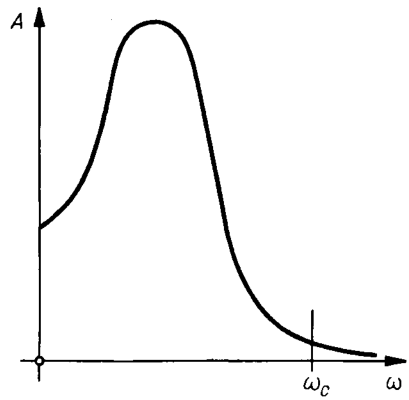


FIG. 185

2. At first sight it would seem that the assumption that the oscillations are nearly harmonic is substantiated also when in the spectrum (5.3) the amplitude A_1^* is immeasurably larger than the other $A_k^*(k = 2, 3, \ldots)$, i. e. when the amplitude characteristic

$$A = \left| \frac{K(i\omega)}{D(i\omega)} \right|$$

of the linear part of the system has a high and sharp peak. But this is not so.

From this supposition alone, however high the peak, it still does not follow that the shape of the oscillations is nearly sinusoidal. This is seen from the following very simple example:

$$\frac{d^2 x}{dt^2} + [rx - f(x)] = 0$$

or

$$\frac{d^2 x}{dt^2} + rx = y, \text{ when } y = f(x).$$

In this case the amplitude characteristic of the linear part of the system $A = \dfrac{1}{r - \omega^2}$ has an infinitely high peak at $\omega = \sqrt{r}$, but the oscillations are nearly harmonic only if the function $f(x)$ is small in modulus. Without this extra condition oscillations very different from harmonic could exist.

Returning to the general case, let us suppose that the function $f(x)$ is nearly linear and can be represented in the form

$$f(x) = rx + \mu\varphi(x),$$

where μ is small. We can attribute the term rx to the linear part of the system. The amplitude characteristic of this "conditional linear system" will be

$$A^* = \left| \frac{K(iw)}{D(i\omega) - rK(i\omega)} \right| .$$

In order that the oscillations shall be nearly harmonic, two conditions must be simultaneously satisfied:

(a) The value of r must be such that the amplitude characteristic of the "conditional linear system" A^* has a sharp and high peak (Fig. 186). This same condition can be formulated as follows: the roots of $D(p) - rK(p) = 0$ must be so placed on the left of the imaginary axis that one pair of roots is near it.

(b) The number μ must be small.

In those cases when these two conditions are satisfied simultaneously, we can say that the system satisfies the conditions of the *auto-resonance hypothesis.*

The conditions of the filter hypothesis and of the auto-resonance hypothesis enable us equally to postulate that the oscillations are nearly harmonic and to find the oscillations in the form (5.4).

But these hypotheses are not equivalent. The following principal differences exist between them:

1. They rest upon different physical properties of the system: the filter hypothesis on the presence in any real system of a finite

range of allowable frequencies, while the auto-resonance hypothesis rests on the properties of several non-linearities weakly propagating harmonics and on the resonance properties of the linear part of the system.

2. The requirements that the function $f(x)$ shall be little different from linear have a different significance in the two cases. With auto-resonance this requirement is necessary even in the exact case (of an infinitely large peak in the amplitude characteristic A^*). For a filter, this requirement is not essential in the exact case (of an ideal filter),

$$A^* = \left| \frac{K(i\omega)}{D(i\omega) - rK(i\omega)} \right|$$

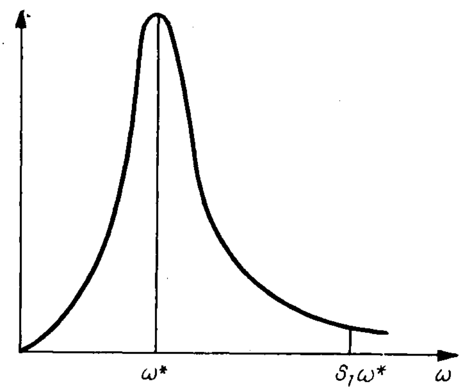


FIG. 186

and arises only when estimating how far the filter can deviate from the ideal.

In this sense the conditions of the filter hypothesis do not depend on the shape of the non-linear characteristic $y = f(x)$. On the contrary, only exception at characteristics can be represented in the form

$$f(x) = rx + \mu\varphi(x)\ ,$$

since the choice of the number r is predetermined by the linear part of the system, and μ must be small.

3. In the case of auto-resonance, we can postulate beforehand that the frequency of the required harmonic solution shall be equal to the frequency corresponding to the peak of the amplitude characteristic A^*, or shall differ little from it. In the case of a filter, it is

not possible to indicate beforehand the value of the required frequency; it can be any of the frequencies lying within the admissible range of the linear part of the system. In this case we verify whether the harmonics have been filtered out, and whether the assumption that the oscillations are nearly harmonic is true only after the assumption has been made and the frequency of oscillation has been determined from it.

In this reasoning we have considered the simplest case, when the system contains one non-linear function of a single argument. But the conclusions we have reached are also true in more general cases. The assumption that the oscillations of any coordinate are nearly harmonic can be substantiated in two cases:

(1) *if the linear parts of the system filter out other frequencies arising from the propagation of oscillations through the non-linear elements, or*

(2) *if the non-linear characteristic is nearly linear, and, in addition, that due to this linear characteristic the amplitude characteristic of the linear part of the system has a high and sharp peak.*

Methods for finding periodic solutions of the equations of the control process in the form (5.4) basically depend on the fact that we are justified by the filter hypothesis or by the auto-resonance hypothesis in supposing that the periodic motions are of this form.

When the harmonic character of the oscillations is explained by a filter, we must look for the solution in the form (5.4), where both the amplitude A and the frequency ω are required. About the frequency, we know only that it must satisfy the inequality

$$\frac{\omega_c}{S} < \omega < \omega_c .$$

In the case of auto-resonance, it is known beforehand that the required frequency is equal to, or little different from, the frequency ω^* for which the amplitude characteristic of the "conditional linear system" has a peak. In this case in equation (5.4) we must take

$$\omega = \omega^* + \delta\omega ,$$

where $\delta\omega$ is a correction to the frequency (of the same order of smallness as μ) and the solution must be sought in the form

$$x = A \sin (\omega^* + \delta\omega)\, t .$$

Methods in which the solution is looked for in the form

$$x = A \sin \omega t ,$$

where A and ω are the unknowns, are therefore called *filter methods.*

An example of a filter method is the method of harmonic balance, of Krylov and Bogolyubov.*

Methods in which the required solution is of the form

$$x = A \sin (\omega^* + \delta\omega) t,$$

where A and $\delta\omega$ are unknowns ($\delta\omega$ is a small number of the order of μ), and where ω^* is known from Fig. 186, are called *auto-resonance methods*. They include such methods as those of Van der Pol, Bulgakov; it is easy to show that the Poincaré method is related to them (if we restrict ourselves to finding the periodic solution to the first approximation) and so are other methods.

In problems of automatic control theory (in contrast to applied electronics problems) the conditions of the filter hypothesis are often realized, while those of the auto-resonance hypothesis are extremely rarely satisfied. We shall therefore give basic attention to filter methods. In Section 6 we demonstrate the process of calculation which is characteristic of auto-resonance methods. In this section we discover for what cases the autoresonance methods lead to the same result as the filter methods.

3. The Approximate Determination of Auto-oscillations by the Harmonic Balance (Filter) Method

Let us consider the simplest special case, when the equations reduce to the form (5.2). *We assume in addition for the present that the characteristic $f(x)$ is odd, but not necessarily single-valued (in particular, it can contain loops)*. We will look for x in the form

$$x = A \sin \omega t , \tag{5.4}$$

where A and ω are the unknowns.

* This method was originally suggested for systems which satisfied the conditions of the auto-resonance hypothesis, although later on it was found that it is also applicable in the case of a filter.

Putting (5.4) in (5.2) we find:

$$\left[a_0 \frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \ldots + a_n\right] A \sin \omega t =$$
$$= \left[b_0 \frac{\mathrm{d}^m}{\mathrm{d}t^m} + b_1 \frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}} + \ldots + b_m\right] f(A \sin \omega t)\,.$$

We now rewrite this equation in the shortened form:

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) A \sin \omega t = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f(A \sin \omega t), \tag{5.5}$$

where

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) = a_0 \frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \ldots + a_n$$

and

$$K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) = b_0 \frac{\mathrm{d}^m}{\mathrm{d}t^n} + b_1 \frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}} + \ldots + b_m$$

are operators, which indicate the operations which must be performed on the function written after them. We expand the periodic function $f[A \sin \omega t]$ in a Fourier series. Since $f(x)$ was assumed to be odd, this series does not contain a free term*:

$$f[A \sin \omega t] = B_1(A) \sin \omega t + C_1(A) \cos \omega t + \ldots$$

We assume that the conditions for the filter hypothesis are satisfied. From the reasoning given in the previous section, it follows that this means we can ignore the harmonics contained in the spectrum of the oscillations of x. Because of this we can omit the higher harmonics in the expansion in series of $f[A \sin \omega t]$, since they give rise to lines in the spectrum of the oscillations of the coordinate which are so small that they can be ignored.

We therefore obtain:

$$f(A \sin \omega t) = B_1 \sin \omega t + C_1 \cos \omega t\,,$$

were B_1 and C_1 are determined by the usual formulae for the co-

* If $f(x)$ is single-valued as well as odd, then $C_1(A) \equiv 0$ (for proof, see p.000).

efficients of a Fourier series:

$$B_1 = \frac{1}{\pi} \int_0^{2\pi} f(A \sin z) \sin z \, \mathrm{d}z = B_1(A),$$

$$C_1 = \frac{1}{\pi} \int_0^{2\pi} f(A \sin z) \cos z \, \mathrm{d}z = C_1(A).$$

Now equation (5.5) can be written:

$$AD\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sin \omega t = B_1(A)\, K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sin \omega t + C_1(A)\, K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \cos \omega t.$$

We now introduce two auxiliary identities.

In the obvious identity

$$L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) e^{i\omega t} = L(i\omega)\, e^{i\omega t},$$

where $L\left(\frac{d}{dt}\right)$ is one of the polynomials $D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ and $K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$, we separate real and imaginary parts:

$$L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) [\cos \omega t + i \sin \omega t] =$$

$$= [\mathrm{Re}\, L(i\omega) + i\, \mathrm{Im}\, L(i\omega)] [\cos \omega t + i \sin \omega t] =$$
$$= [\mathrm{Re}\, L(i\omega) \cos \omega t - \mathrm{Im}\, L(i\omega) \sin \omega t] +$$
$$+ i [\mathrm{Re}\, L(i\omega) \sin \omega t + \mathrm{Im}\, L(i\omega) \cos \omega t].$$

Equating them separately to zero, we obtain:

$$L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sin \omega t = \mathrm{Im}\, L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) e^{i\omega t} =$$
$$= \mathrm{Re}\, L(i\omega) \sin(\omega t) + \mathrm{Im}\, L(i\omega) \cos \omega t,$$

$$L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \cos \omega t = \mathrm{Re}\, L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) e^{i\omega t} =$$
$$= \mathrm{Re}\, L(i\omega) \cos wt - \mathrm{Im}\, L(i\omega) \sin \omega t.$$

Taking in turn $L\left(\frac{d}{dt}\right)=D\left(\frac{d}{dt}\right)$ and $-L\left(\frac{d}{dt}\right)=K\left(\frac{d}{dt}\right)$ and using the above identities, we perform the operations

$$D\left(\frac{d}{dt}\right)\sin\omega t,\ K\left(\frac{d}{dt}\right)\sin\omega t$$

and

$$K\left(\frac{d}{dt}\right)\cos\omega t.$$

The equation (5.5) then reduces to the form

$$A[\operatorname{Re}D(i\omega)\sin\omega t+\operatorname{Im}D(i\omega)\cos\omega t]=$$
$$=B_1(A)\,[\operatorname{Re}K(i\omega)\sin\omega t+\operatorname{Im}K(i\omega)\cos\omega t]+$$
$$+C_1(A)\,[\operatorname{Re}K(i\omega)\cos\omega t-\operatorname{Im}K(i\omega)\sin\omega t].$$

In this equation we equate the coefficients of $\sin\omega t$ and $\cos\omega t$ separately to zero and find

$$A\operatorname{Re}D(i\omega)=B_1(A)\operatorname{Re}K(i\omega)-C_1(A)\operatorname{Im}K(i\omega),$$
$$A\operatorname{Im}D(i\omega)=B_1(A)\operatorname{Im}K(i\omega)+C_1(A)\operatorname{Re}K(i\omega).$$

Multiplying the second equation by i and adding it to the first

$$AD(i\omega)=B_1(A)\,K(i\omega)+C_1(A)\,[i\operatorname{Re}K(i\omega)-\operatorname{Im}K(i\omega)].$$

If we now bring i out of the square brackets, we obtain:

$$AD(i\omega)=[B_1(A)+iC_1(A)]\,K(i\omega),$$

or

$$\boxed{I(i\omega)=R(A),} \tag{5.6}$$

where

$$I(i\omega)=\frac{D(i\omega)}{K(i\omega)};\quad R(A)=\frac{B_1(A)+iC_1(A)}{A}.$$

The expression on the left-hand side of equation (5.6)

$$I(i\omega)=\frac{D(i\omega)}{K(i\omega)}$$

itself represents the amplitude-phase characteristic of the first kind of the linear part of the system, i.e. the ratio of the complex amplitude of the input quantity of this part of the system to the complex amplitude of its output coordinate for steady oscillations.

The expression on the right.hand side of equation (5.6)

$$R(A) = \frac{B_1(A) + iC_1(A)}{A} = B(A) + iC(A) \tag{5.7}$$

is equivalent to the amplitude-phase characteristic of the second kind for a non-linear element. An oscillation of amplitude A acts at the input of the non-linear element, and at its output an oscillation having a first harmonic with complex amplitude $B_1 + iC_1$ is set up. Hence $R(A)$ is the ratio of the complex amplitude of the first harmonic of the oscillation set up at the output of the non-linear element to the amplitude of the steady oscillation at its input. The quantity $R(A)$ is sometimes called the *reduced transfer function of the non-linear element*, or the *describing function*.

Functions $R(A)$ for some typical non-linearities are set out in Table XVI.

In the complex plane we now construct two hodographs (Fig. 187): the hodograph $I(i\omega)$, i.e. the amplitude-phase characteristic of the first kind of the linear part of the system (the parameter being the required frequency ω), and the hodograph $R(A)$, the reduced transfer function of the non-linear element (the parameter being the required amplitude A). *On the strength of equation (5.6) the point of intersection of these two hodographs determines the required amplitude A and the frequency of the periodic solution.*

Of course, the results are true only if the initial assumption, that the periodic solutions of the considered equations were nearly harmonic, is true. The validity of this assumption can now be tested from the amplitude characteristic in the way explained at the end of the previous section. We can also do this directly from the amplitude-phase characteristic of the first kind which was constructed to prove the existence of periodic solutions. For a filter the radius-vectors of the points corresponding to the frequencies of the harmonics are considerably larger than the radius-vector of the point corresponding to the obtained frequency of the periodic solution.

The points of intersection of the hodographs in Fig. 187 for which the frequencies do not satisfy this condition must be excluded from our consideration.

If the hodographs $I(i\omega)$ and $R(A)$ intersect at several points, and each of them satisfies this condition which results from the filter hypothesis, this points to the presence of several periodic solutions: each point of intersection of the hodograph determines one of them.

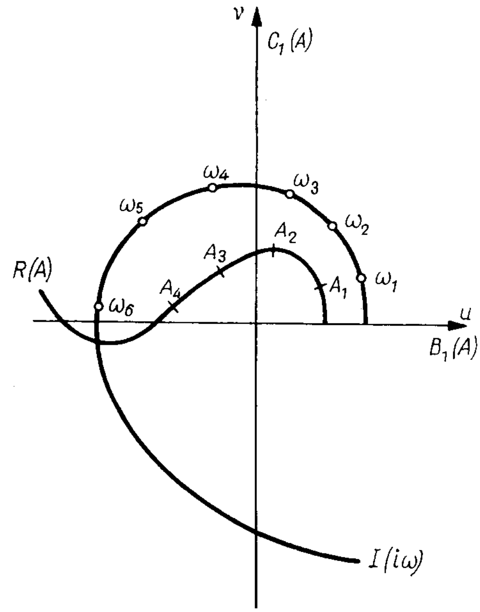


FIG. 187

If there are no points of intersection, there is no nearly harmonic periodic solution. From this, of course, it does not follow that the system in general does not have periodic solutions, since by using filter methods we can only find periodic solutions whose frequencies satisfy the inequality

$$\frac{\omega_c}{S} < \omega < \omega_c ,$$

and the question of the existence of periodic solutions with $\omega < \dfrac{\omega_c}{S}$ remains unanswered, since they cannot be found in the form (5.4).

We assumed above that the characteristic $f(x)$ was odd. This assumption is rarely fulfilled. Indeed, in static systems, the position

TABLE XVI

FORMULAE FOR THE CALCULATION OF R(A) FOR TYPICAL NON-LINEARITIES

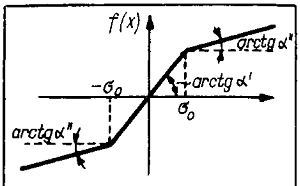
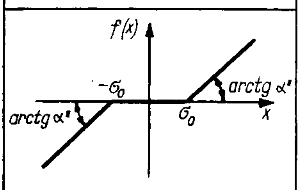
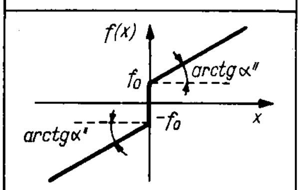
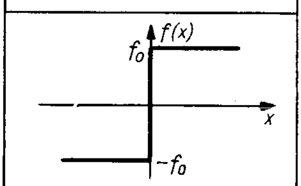
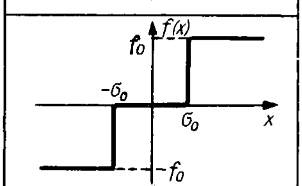
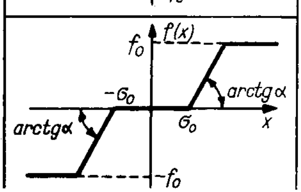
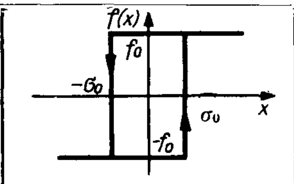
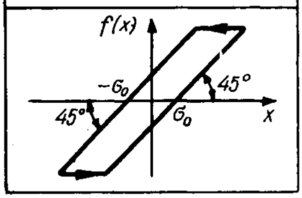
<table>
<tr><th>No.</th><th>Graph of (x)</th><th>$R(A)$</th></tr>
<tr><td>I</td><td>
</td><td>$$B(A) = a'' + \frac{2}{\pi}(a' - a'')\left[\arcsin\frac{\sigma_0}{A} + \frac{\sigma_0}{A}\sqrt{1-\left(\frac{\sigma_0}{A}\right)^2}\right],\ C(A) = 0$$</td></tr>
<tr><td>II</td><td>
</td><td>$$B(A) = a'' - \frac{2}{\pi}\left[\arcsin\frac{\sigma_0}{A} + \frac{\sigma_0}{A}\sqrt{1-\left(\frac{\sigma_0}{A}\right)^2}\right].\ C(A) = 0$$</td></tr>
<tr><td>III</td><td>
</td><td>$$B(A) = a'' + \frac{4f_0}{\pi A},\ C(A) = 0$$</td></tr>
<tr><td>IV</td><td>
</td><td>$$B(A) = \frac{4f_0}{\pi A},\ C(A) = 0$$</td></tr>
<tr><td>V</td><td>
</td><td>$$B(A) = \frac{4f_0}{\pi A}\sqrt{1-\left(\frac{\sigma_0}{A}\right)^2},\ C(A) = 0$$</td></tr>
<tr><td>VI</td><td>
</td><td>For $A > \sigma_0 + \frac{f_0}{2}$, $B(A) = \frac{2a}{\pi}(\sigma_2 - \sigma_1 + \sin 2\sigma_2 - \sin\sigma_1$ where $\sigma_1 = \arcsin\frac{\sigma_0}{A} < \frac{\pi}{2}$, $\sigma_2 = \arcsin\frac{1}{A}\left(\sigma_0 + \frac{f_0}{a}\right)\frac{\pi}{2}$, $C(A) = 0$</td></tr>
</table>

TABLE XVI

FORMULAE FOR THE CALCULATION OF R(A) FOR TYPICAL NON-LINEARITIES

No.	Graph of (x)	$R(A)$
VII		$B(A) = \frac{4f_0}{\pi A}\sqrt{1-\left(\frac{\sigma_0}{A}\right)}, \; C(A) = \frac{4f_0\sigma_0}{\pi A^2}$
VIII		$B(A) = \frac{1}{\pi}\left(\psi_1 - \frac{1}{3}\sin 2\psi_1\right), \; C(A) = \frac{1}{\pi}\sin^2\psi_1$ where $\psi_1 = \arccos\left[1 - \frac{2(A-\sigma_0)}{A}\right] < \pi$

on the characteristic of the point of equilibrium which determines the origin of x depends on the load, and even if the characteristic is symmetric with respect to some point, only for some unique value of the load does the origin of coordinates turn out to be at this point. Whatever the form of the characteristic $f(x)$, the assumption that it is odd has, therefore, no meaning when the system is static. *For static systems equation (5.6) is not suitable for determining the periodic states.*

But if $f(x)$ is not odd, and, at the same time, the equation $D(p) = 0$ has no zero roots, the existence of periodic states becomes noticeably more complicated.

In this case the periodic solution must be looked for in the form

$$x = E + A \sin \omega t ,$$

where $0 < E < \infty$, and it is necessary to determine the three quantities E, A and ω.

Putting this value of x in the equation

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f(x),$$

we obtain:

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)[E + A\sin\omega t] = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f[E + A\sin\omega t].$$

The first terms of the expansion of the function $f[E + A \sin \omega t]$ in a Fourier series are equal to

$f(E + A \sin \omega t) \approx S(E, A) + B_1(E, A) \sin \omega t + C_1(E, A) \cos \omega t.$

Then, just as before, in determining the periodic states, we obtain two equations

$$\boxed{I(i\omega) = R(E, A)} \tag{5.8}$$

and

$$\boxed{a_n E = b_n S(E, A),} \tag{5.8'}$$

where

$$R(E, A) = \frac{B_1(E,A) + iC_1(E, A)}{A}.$$

The equation (5.8), as before, is obtained by equating the terms containing $\sin \omega t$ and $\cos \omega t$, and the equation (5.8′) by equating the free terms.

It is immediately seen that for both the conditions $a_n = 0$ (i.e. $E = \infty$) and the condition $b_n = 0$ (i.e. $E = 0$) it is not possible to find a periodic solution in the form

$$x = E + A \sin \omega t, \; 0 < E < \infty.$$

The case $b_n = 0$ is not encountered in control problems, since this would mean that the loop of static actions was closed. In the case $a_n = 0$ the system contains an astatic stage, and such a system cannot be closed by a non-linear stage with a non-symmetric (not odd) characteristic, since an astatic stage, by integrating the static deviation, continuously increases it and $E \to \infty$.

If $a_n \neq 0$ and $b_n \neq 0$, then it is necessary first of all to determine all pairs of numbers E, A, satisfying (5.8′). To do this we construct in the F, E-plane the family of curves $F = S(E, A)$ for various values of A, and find the points of intersection of these curves with the ray $F = \frac{a_n}{b_n} E$ (Fig. 188).

The pairs of numbers E, A, satisfying equation (5.8′) which we obtain are substituted in $R(E, A)$. In the construction of this hodo-

graph, each of its points bears not one, but two values. Its intersection with the hodograph $I(i\,\omega)$ determines at once the three required numbers: the amplitude of the auto-oscillations, A, their frequency ω and the magnitude of the displacement of the zero line of E (Fig. 189). Of course, each time it is necessary to verify that the frequencies of the periodic states found satisfy the conditions of the filter hypothesis.

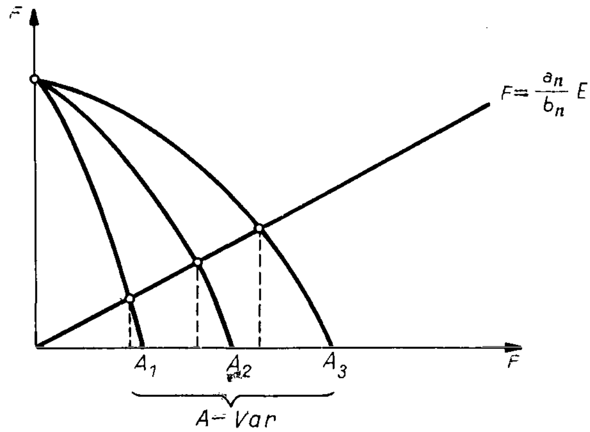


FIG. 188

Returning to odd characteristics we note, in conclusion, that in the case when the function $f(x)$ is not only odd, but also single-valued, the coefficient $C_1 = 0$, and the hodograph $R(A)$ lies along the real axis.

For

$$C_1 = \frac{1}{\pi}\int_0^{2\pi} f(A\sin z)\cos z\,dz = \frac{1}{\pi A}\left[\int_0^{\frac{\pi}{2}} f(A\sin z)\,A\cos z\,dz + \right.$$

$$+ \int_{\frac{\pi}{2}}^{\pi} f(A\sin z)\,A\cos z\,dz + \int_{\pi}^{\frac{3}{2}\pi} f(A\sin z)\,A\cos z\,dz +$$

$$\left. + \int_{\frac{3}{2}\pi}^{2\pi} f(A\sin z)\,A\cos z\,dz\right].$$

We perform the substitution

$$\psi = A \sin z ,$$

giving

$$\mathrm{d}\psi = A \cos z \, \mathrm{d}z .$$

Then

$$C_1 = \frac{1}{\pi A} \left[\int_0^A f(\psi)\, \mathrm{d}\psi + \int_A^0 f(\psi)\, \mathrm{d}\psi + \right.$$

$$\left. + \int_0^{-A} f(\psi)\, \mathrm{d}\psi + \int_{-A}^0 f(\psi)\, \mathrm{d}\psi \right]$$

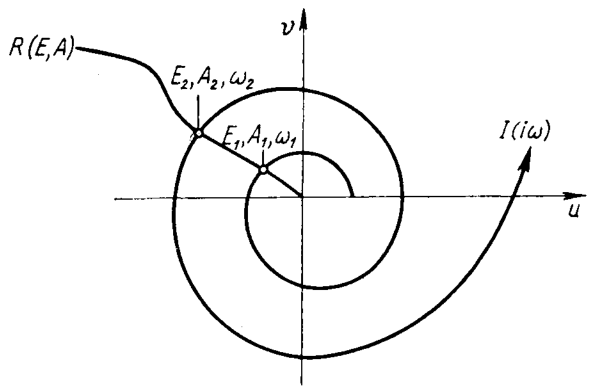


FIG. 189

If the characteristic $f(x)$ is not single-valued, i.e. it contains a loop, and if the area of this loop between $x = -A$ and $x = A$ is equal to F, the sum of the four integrals in square brackets, is F.

Therefore,

$$C_1 = \frac{F}{\pi A} .$$

In the special case of a single-valued characteristic $F = 0$, and therefore $C_1 = 0$.

In this case it is not necessary to construct the hodographs and the whole construction can be performed in the real plane. Thus, for an odd characteristic ($E = 0$) from

$$I(i\omega) = R(A)$$

for real $R(A)$ we obtain

$$\operatorname{Im} I(i\omega) = 0\,,$$

$$\operatorname{Re} I(i\omega) = R(A)\,.$$

From this it follows that when $f(x)$ is a single-valued function (i.e. does not contain a loop) the frequencies are determined independently of the amplitudes and depend only on the linear part of the system. They are equal to the complex roots of the coefficient of the imaginary part of

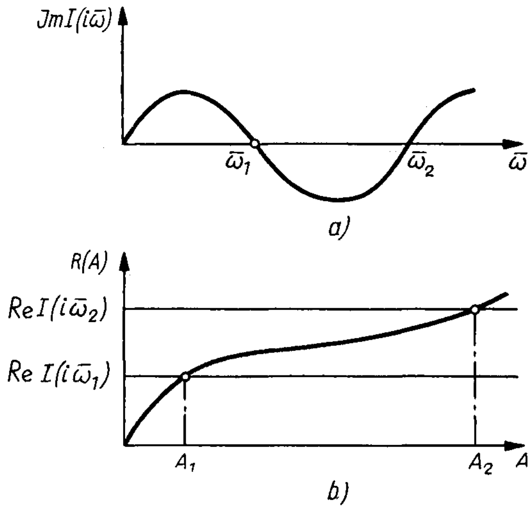


FIG. 190

the characteristic equation of the conditional linear system, which would be obtained if the given system were closed through a linear inertialess amplifier $y = kx$ instead of a non-linear element, and if k were then increased until a pair of roots of the characteristic equation reached the imaginary axis ($k = k_{cr}$). Therefore, for single-valued characteristics the frequencies of the nearly harmonic periodic solutions do not depend on the form of the characteristic of the non-linear element and are found by the intersection of the curve

$$R = R(A)$$

with the straight line

$$R = \operatorname{Re} I(i\,\bar{\omega})$$

(Fig. 190*b*).

In fact, such a conditional system has the characteristic equation

$$D(p) - kK(p) = 0.$$

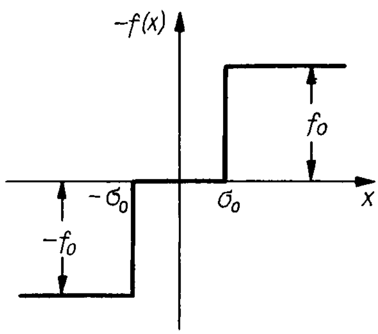


Fig. 191

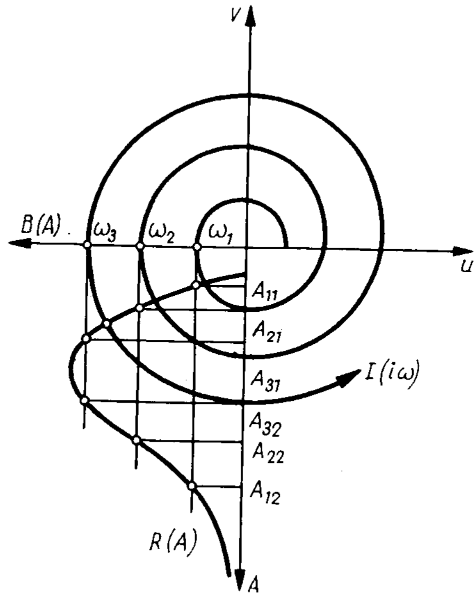


Fig. 192

The boundary of the D-partition for k is defined as

$$k = \frac{D(i\omega)}{K(i\omega)} = I(i\omega).$$

The boundary values of k are determined by the condition

$$k_{cr} = \operatorname{Re} I(i\bar{\omega}),$$

where $\bar{\omega}$ is found (Fig. 190a) from $\operatorname{Im} I(i\bar{\omega}) = 0$, i. e. from the same conditions which gave the frequency of the auto-oscillations for a single-valued non-linear characteristic.

The amplitude of the auto-oscillations in the given case is determined simply from the condition

$$R(A) = k_{cr}.$$

EXAMPLE. A single-loop network is closed through a relay with "dead-space" (Fig. 191).

The equations of the non-linear element are:

$$y = f(x) = \begin{cases} -f_0 & \text{when} \quad x > \sigma_0, \\ 0 & \text{when} \; -\sigma_0 < x < \sigma_0, \\ +f_0 & \text{when} \quad x < -\sigma_0. \end{cases} \tag{5.9}$$

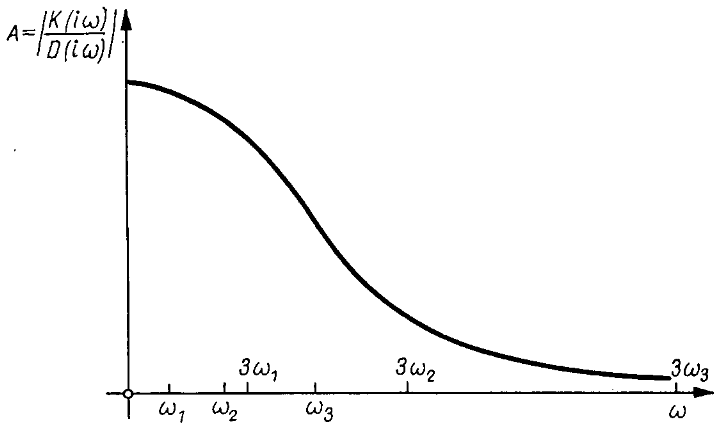


FIG. 193

For such a relay (see Table XVI)

$$B(A) = \frac{4f_0}{\pi A}\sqrt{1-\left(\frac{\sigma_0}{A}\right)^2},$$

$$C(A) = 0.$$

Constructing the hodographs $I(i\,\omega)$ and $R(A) = B(A)$ on one graph, we find from it the frequencies and amplitudes of the possible periodic states (Fig. 192). Of course, the states we obtain can include unstable states too (see below).

We construct the amplitude characteristic of the linear part of the system (Fig. 193) and along the ω-axis we put the obtained frequencies ω_1, ω_2 and ω_3 and the frequencies of the harmonics* $3\omega_1$, $3\omega_2$, and $3\omega_3$. Suppose, for example, that they are situated along the ω-axis as in Fig. 193. The frequencies $3\omega_2$ and $3\omega_3$ lie outside the admissible range of the system, and therefore the system with equation (5.9) has a periodic solution which is nearly harmonic with frequencies ω_2 and ω_3 and with the corresponding amplitudes A_2 and A_3.

The periodic solution we have found with frequency ω_1 is doubtful and must be verified by an exact method, since it has been found on the assumption that the oscillations are nearly harmonic, and the frequency lies within the admissible range of the linear part of the system.

4. The Approximate Determination of Forced Oscillations in the Presence of an External Periodic Action

So far in this chapter we have spoken of an automatic control system containing non-linear elements, but free from any external actions. Often it is necessary to consider non-linear systems subject to external periodic actions.

The motions arising in a non-linear system under the action of external periodic actions differ essentially from those arising in linear systems with similar conditions.

We recall the three basic properties of the steady motions possible in linear systems under the action of external periodic disturbances:

(a) The steady motions are always periodic; their frequency is the same as the frequency of the external action.

(b) The amplitude of the steady motion is directly proportional to the amplitude of the external action and is a single-valued function of the frequency.

(c) The stability of the steady motions depends only on the properties of the system, and is completely independent of the amplitude and frequency of the external action.

From this, as we showed in Chapter II, a single curve serves as the amplitude-phase characteristic of a linear system.

In contrast to a linear system, in non-linear systems:

(a) As well as frequencies and amplitudes of the external action for which the steady motion is periodic with the same frequency, there can exist frequencies and amplitudes of external actions which cause non-periodic (almost periodic) steady motion. In this case this steady motion is the sum of several harmonic oscillations. One of them has a frequency coinciding with that of the external disturbance, and the second (or others) has a frequency near to that of the auto-oscillation (or auto-oscillations) which exist or are possible in the system when the external periodic action is absent.

(b) The amplitude of the periodic steady motions and the amplitudes of the separate periodic constituents of the non-periodic steady motions are not proportional to the amplitude of the external action,

* The frequencies $2\omega_1$, $2\omega_2$ and $2\omega_3$ are not in the spectrum, since the characteristic is odd.

but are non-linear functions of it. The form of this function depends on the properties of the system.

The amplitude of the steady motions also need not be a single-valued function of the frequency and amplitude of the external action.

(c) The stability of both periodic and non-periodic steady motions depends not only on the properties of the system, but also on the amplitude and frequency of the external action.

This means that the amplitude, or amplitude-phase, characteristic of a non-linear system, determined by analogy with the same characteristic of a linear system, cannot be constructed for all values of the frequencies and amplitudes of the external action, but only for those which cause periodic steady motions in the system. Moreover, such characteristics consist not of a single curve, but of a single-parameter family of curves: each curve corresponds to one value of the amplitude of the external action.

Finally, some sections of these curves can correspond to stable, and others to unstable, periodic motions.

We describe below the application of the harmonic balance method to the approximate determination of similar characteristics for non-linear systems.

Suppose that at the input of the linear part of the system (see Fig. 180) an external periodic action S sin ωt, acts in addition to the output coordinate of the non-linear element, i.e. the equation of motion is of the form

$$\left.\begin{aligned} D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x &= K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)[y + S\sin\omega t],\\ y &= f(x) \end{aligned}\right\} \tag{5.10}$$

or

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)[f(x) + S\sin\omega t]. \tag{5.11}$$

Here $D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ and $K\left(\frac{\mathrm{d}t}{\mathrm{d}}\right)$ are the operators whose meaning was explained in the previous section.

We restrict ourselves to the case when the function $f(x)$ is odd, and, consequently, the oscillations of the coordinate can only be symmetric with respect to $x = 0$.

We discussed above the fact that in auto-oscillatory or potentially auto-oscillatory systems, when an external periodic disturbance is present, both periodic and almost periodic oscillations can be set up. Here we only consider the question of determining the periodic steady states. Now it is no longer possible to arbitrarily dispose of the time coordinate origin — it is fixed since t enters explicitly in equation (5.10). Consequently, in the given case it is essential to determine not only the amplitude of the steady oscillations, but also their phase shift relative to the external harmonic disturbance. Therefore we look for the steady periodic oscillations of the coordinate x in the form

$$x = A \sin(\omega t - \gamma)\,, \tag{5.12}$$

where A and γ are constants to be determined, and ω is the frequency coinciding with the given frequency of the external disturbance.

Putting (5.12) in (5.11) we obtain:

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) A \sin(\omega t - \gamma) = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\{f[A \sin(\omega t - \gamma)] + S \sin \omega t\}. \tag{5.13}$$

As in the previous section, we replace the periodic function $f[A \sin(\omega t - \gamma)]$ by the first harmonics of its expansion in a Fourier series:

$$f[A \sin(\omega t - \gamma)] \approx B_1(A) \sin(\omega t - \gamma) + C_1(A) \cos(\omega t - \gamma)\,. \tag{5.14}$$

Making use of the identities

$$\left.\begin{aligned} L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sin \omega t &= \operatorname{Re} L(i\omega) \sin \omega t + \operatorname{Im} L(i\omega) \cos \omega t,\\ L\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \cos \omega t &= \operatorname{Re} L(i\omega) \cos \omega t - \operatorname{Im} L(i\omega) \sin \omega t, \end{aligned}\right\} \tag{5.15}$$

which were proved in the previous section, and from the obvious identity:

$$\sin \omega t = \sin(\omega t - \gamma) \cos \gamma + \cos(\omega t - \gamma) \sin \gamma\,,$$

we can reduce the equation (5.13) to the form

$$
\begin{aligned}
&A[\operatorname{Re} D(i\,\omega) \sin(\omega\, t - \gamma) + \operatorname{Im} D(i\,\omega) \cos(\omega\, t - \gamma)] = \\
&= B_1(A)\, [\operatorname{Re} K(i\,\omega) \sin(\omega\, t - \gamma) + \operatorname{Im} K(i\,\omega) \cos(\omega\, t - \gamma)] + \\
&\quad + C_1(A)\, [\operatorname{Re} K(i\,\omega) \cos(\omega\, t - \gamma) - \operatorname{Im} K(i\,\omega) \sin(\omega\, t - \gamma)] + \\
&\quad + S \cos\gamma [\operatorname{Re} K(i\,\omega) \sin(\omega\, t - \gamma) + \operatorname{Im} K(i\,\omega) \cos(\omega\, t - \gamma)] + \\
&\quad + S \sin\gamma [\operatorname{Re} K(i\,\omega) \cos(\omega\, t - \gamma) - \operatorname{Im} K(i\,\omega) \sin(\omega\, t - \gamma)].
\end{aligned}
\tag{5.16}
$$

In this equation we equate the coefficients of the terms in $\sin(\omega\, t - \gamma)$ and the coefficients of the terms in $\cos(\omega\, t - \gamma)$:

$$
\left.
\begin{aligned}
A \operatorname{Re} D(i\,\omega) &= B_1(A) \operatorname{Re} K(i\,\omega) - C_1(A) \operatorname{Im} K(i\,\omega) + \\
&\qquad + S \cos\gamma \operatorname{Re} K(i\,\omega) - S \sin\gamma \operatorname{Im} K(i\,\omega)\,, \\
A \operatorname{Im} D(i\,\omega) &= B_1(A) \operatorname{Im} K(i\,\omega) + C_1(A) \operatorname{Re} K(i\,\omega) + \\
&\qquad + S \cos\gamma \operatorname{Im} K(i\,\omega) + S \sin\gamma \operatorname{Re} K(i\,\omega)\,.
\end{aligned}
\right\}
\tag{5.17}
$$

Multiplying now the second equation of this system by i and adding it to the first, we obtain:

$$
AD(i\omega) = [B_1(A) + iC_1(A)]\, K(i\omega) + SK(i\omega)\ (\cos\gamma + i \sin\gamma).
\tag{5.18}
$$

As before, we put

$$
R\,(A) = \frac{1}{A}\, [B_1\,(A) + iC_1(A)].
$$

Then, dividing both sides of the equation (5.18) by $AK(i\omega)$, we obtain finally

$$
\boxed{I\,(i\omega) = R\,(A) + \frac{S}{A}\, e^{i\gamma}.}
\tag{5.19}
$$

In this vector equation S and ω are given and A and γ are to be determined. Let us assume, however, that S and A are given, and determine

ω and γ for which equation (5.19) is satisfied. To do this we return to the construction performed in the previous section to determine the auto-oscillatory state and to the two curves contained in it: the characteristic and the reduced transfer function $R(A)$ (Fig. 194).

We take the value $A = A_1$ and determine the values of ω and γ for which equation (5.19) is satisfied. To do this we put a compass point at the point of the hodograph $R(A)$ corresponding to $A = A_1$, and draw a circle of radius $\dfrac{S}{A_1}$. The point of intersection of this circle

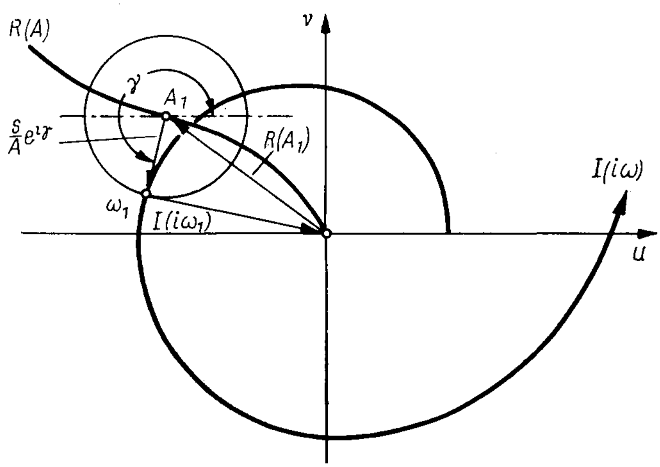


FIG. 194

with the hodograph $I(i\omega)$ determines the required value of ω_1, and the angle, read anticlockwise as in Fig. 194, is the value of γ.

In fact, in this case a vector triangle is formed, determined by (5.19) and shown in Fig. 194.

If the circumference has no points of intersection with the hodograph $I(i\,\omega)$, this indicates that there are no periodic motions with the amplitude $A = A_1$ in the system for any value of ω. If, finally, there are several points of intersection, there also exist several frequencies for which the oscillations of the coordinate x have the amplitude $A = A_1$.

By changing the values of A we can determine all the periodic motions possible in the system for any values of ω or S.

5. Systems Containing Several Non-Linearities

We return to a consideration of systems which are not subject to external periodic disturbances. We consider now systems which contain several *odd single-valued* non-linear functions so that the system can be represented as a sequential network of the linear parts of the system and of the non-linear elements with odd and single-valued characteristics (Fig. 195).

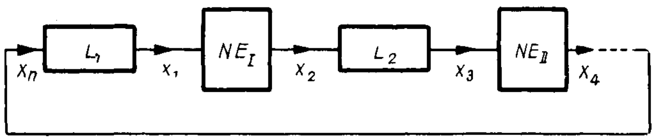


FIG. 195

The equations of such a system are

$$\left.\begin{aligned} D_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_1 &= K_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_n,\\ x_2 &= f_1(x_1),\\ D_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_3 &= K_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_2,\\ x_4 &= f_2(x_3),\\ &\cdots\cdots\cdots\cdots\\ D_{n-2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_{n-1} &= K_{n-2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_{n-2},\\ x_n &= f_{\frac{n}{2}}(x_{n-1}). \end{aligned}\right\}$$

Assuming that the oscillations of any coordinate of the system, for example of x_n, are nearly harmonic, we will look for the periodic solution for x_n in the form

$$x_n = A_n \sin \omega t. \tag{5.20}$$

Then

$$x_1 = \frac{A_n}{\left|\dfrac{D_1(i\omega)}{K_1(i\omega)}\right|}\sin\left[\omega t - \arg\frac{D_1(i\omega)}{K_1(i\omega)}\right] = A_1\sin(\omega t - \varphi_1),$$

where

$$A_1 = \frac{A_n}{\left|\dfrac{D_1(i\omega)}{K_1(i\omega)}\right|}, \qquad \varphi_1 = \arg\frac{D_1(i\omega)}{K_1(i\omega)}.$$

The oscillations of the coordinate x_2 are:

$$x_2 = f_1[A_1 \sin(\omega t - \varphi_1)] = A_1 R_1(A_1)\sin(\omega t - \varphi_1) + \ldots$$

Here $R_1(A_1)$ is the reduced transfer function of the first non-linear element, so that $A_1 R(A_1)$ is the amplitude of the basic mode, and the dots denote harmonics.* Rejecting the harmonics** we obtain:

$$x_3 = \frac{A_1 R_1(A_1)}{\left|\dfrac{D_2(i\omega)}{K_2(i\omega)}\right|}\sin\left|\omega t - \varphi_1 - \arg\frac{D_2(i\omega)}{K_2(i\omega)}\right| = A_3 \sin(\omega t - \varphi_3),$$

where

$$A_3 = \frac{A_1 R_1(A_1)}{\left|\dfrac{D_2(i\omega)}{K_2(i\omega)}\right|}, \qquad \varphi_3 = \varphi_1 + \arg\frac{D_2(i\omega)}{K_2(i\omega)}.$$

Therefore

$$x_4 = f_2[A_3 \sin(\omega t - \varphi_3)] = A_3 R_2(A_3)\sin(\omega t - \varphi_3) + \ldots$$

repeating the argument and moving further along the action network in this way, we obtain:

$$x_{n-1} = \frac{A_{n-3} R_{\frac{n}{2}-1}(A_{n-3})}{\left|\dfrac{D_{\frac{n}{2}}(i\omega)}{K_{\frac{n}{2}}(i\omega)}\right|}\sin\left(\omega t - \varphi_{n-3} - \arg\frac{D_{\frac{n}{2}}(i\omega)}{K_{\frac{n}{2}}(i\omega)}\right) = \tag{5.21}$$

$$= A_{n-1}\sin(\omega t - \varphi_{n-1}),$$

$$x_n = A_{n-1} R_{\frac{n}{2}}(A_{n-1})\sin(\omega t - \varphi_{n-1}).$$

* The phase of the basic mode coincides with the phase of the oscillations of x_1, since the function f_1 is single-valued (see p. 355).

** See the end of this section for the conditions under which this is allowable.

Equating (5.20) and (5.21) we find:

$$A_n = A_{n-1} R_{\frac{n}{2}}(A_{n-1}), \qquad -\varphi_{n-1} = 0. \tag{5.22}$$

Substituting the values of A_{n-1}, and then of A_{n-2} and so on, and dividing by A_n, we obtain:

$$\left|\frac{D_1(i\omega)}{K_1(i\omega)}\right| \cdot \left|\frac{D_2(i\omega)}{K_2(i\omega)}\right| \cdots \left|\frac{D_{n-2}(i\omega)}{K_{n-2}(i\omega)}\right| =$$

$$= R_1(A_1) R_2\left[\left|\frac{K_2(i\omega)}{D_2(i\omega)}\right| A_1 R_1(A_1)\right] \times$$

$$\times R_3\left\{\left|\frac{K_2(i\omega)}{D_2(i\omega)}\right| \left|\frac{K_3(i\omega)}{D_3(i\omega)}\right| R_2\left[\left|\frac{K_2(i\omega)}{D_2(i\omega)}\right| A_1 R_1(A_1)\right]\right\} \cdots$$

We denote the product of the functions on the right-hand side of the equation $R(A_1, \omega)$ and call it the *free reduced transfer function of all the non-linearities.*

We note now that the left-hand side of the resulting equation itself represents the amplitude characteristic of the linear system which would be obtained from the given system if all the non-linear elements were removed and all the linear elements in the sequential network were connected. We denote it by

$$\left|\frac{D(i\omega)}{K(i\omega)}\right| = |I(i\omega)|.$$

Then

$$\boxed{|I(i\omega)| = R(\overline{A_1, \omega}).} \tag{5.23}$$

We obtain a second relation from the second equation of (5.22). It is

$$-\left[\arg\frac{D_1(i\omega)}{K_1(i\omega)} + \arg\frac{D_2(i\omega)}{K_2(i\omega)} + \ldots\right] = 0$$

or

$$\boxed{-\arg I(i\omega) = 0.} \tag{5.24}$$

This condition does not depend on the amplitude and at once enables

us to determine ω. It is immediately seen that *the frequencies of the auto-oscillations coincide with the frequencies on the D-partition boundary for the coefficient of amplification in the system obtained from the given one by replacing the non-linear elements by linear inertialess amplifiers.* Hence, the frequency does not depend on the number of single-valued non-linearities or their form, and depends only on the linear parts of the system.*

Putting the frequencies found from (5.24) in (5.23) we find the amplitude A_1 for each frequency.

From the reasoning given it follows that it is true only if the system can be reduced to the form shown in Fig. 195, if all the non-linear characteristics are odd and single-valued, and if the harmonics produced by the non-linearities can be ignored. This last condition is satisfied only if the linear sections of the system *separately* block frequencies of $3\omega^*$ and higher, where the ω^* are the frequencies found from (5.24). This condition must be checked after ω^* has been determined from all the amplitude characteristics $\left|\frac{D_j(i\omega)}{K_j(i\omega)}\right|$ separately.

If the first of these conditions is not satisfied, then similar reasoning leads to two related equations for the determination of ω and A_1, whose solution is considerably more complicated. It requires the construction of a family of hodographs.

6. The Determination of the Auto-oscillations in the Case of Auto-resonance

In Sections 3, 4 and 5 it was assumed that the system satisfied the filter hypothesis, and the auto-oscillatory states were determined by the harmonic balance method (by the filter method), i.e. we looked for the solution in the form

$$x = A \sin \omega t ,$$

where ω could be any number which statisfied the inequality:

$$\frac{\omega_c}{S} < \omega < \omega_c .$$

* Cf. p. 355.

Let us now suppose that the required periodic solutions can be assumed to be nearly harmonic not because of a filter but because of auto-resonance, i.e. that the equations are of the form

$$\left.\begin{aligned} D(p)\,x &= K(p)\,y\,, \\ y = f(x) &= rx + \mu\varphi(x)\,, \end{aligned}\right\} \tag{5.25}$$

where μ is small, and r is a positive number such that the characteristic equation

$$D(p) - rK(p) = 0 \tag{5.26}$$

of the "conditional linear system" obtained from (5.25) for $\mu = 0$ has a pair of imaginary roots $p_{1,2} = \pm i\,\omega^*$, and its other roots lie to the left of the imaginary axis.

Let us suppose that $f(x)$ is an odd function. We shall now look for periodic solutions in the form

$$x = A \sin(\omega^* + \delta\omega)\,t\,, \tag{5.27}$$

where $\delta\omega$ is a small "correction to the frequency"; it is of the same order as μ.

Putting for the time being

$$\omega = \omega^* + \delta\omega\,, \tag{5.28}$$

we put (5.27) in (5.25), and repeating the calculations performed in Section 3 above we obtain the equation (5.6). But now, taking (5.28) nito account, we can rewrite this equation as

$$I[(\omega^* + \delta\omega)] = R(A)$$

or

$$D[i(\omega^* + \delta\omega)] = R(A)\,K[i(\omega^* + \delta\omega)]. \tag{5.29}$$

Expanding the terms $D[i(\omega^* + \delta\omega)]$ and $K[i(\omega^* + \delta\omega)]$ in series of ascending powers of $i\,\delta\omega$ and ingoring the terms containing the small quantity $\delta\omega$ to the second and higher degrees, we find

$$\begin{aligned} D(i\omega^*) + \left[\frac{\mathrm{d}D(i\omega)}{\mathrm{d}(i\omega)}\right]_{\omega=\omega^*} i\delta\omega - \Bigg\{ K(i\omega^*) &+ \\ + \left[\frac{\mathrm{d}K(i\omega)}{\mathrm{d}(i\omega)}\right]_{\omega=\omega^*} &i\delta\omega \Bigg\} R(A) = 0. \end{aligned} \tag{5.30}$$

We recall that in the given case, due to autoresonance,

$$f(x) = rx + \mu\varphi(x)$$

and therefore

$$R(A) = r + \mu R^*(A) , \tag{5.31}$$

where R^* is the aggregate of terms containing μ as a factor.

Putting (5.31) in (5.30) and ignoring terms which contain the product of the small quantities μ and $\delta\omega$, since this product is of the second order, we obtain

$$D(i\omega^*) - R(A)\, K(i\omega^*) = -i\, \Delta'(i\omega^*)\, \delta\omega , \tag{5.32}$$

where

$$\Delta\,(i\omega) = D\,(i\omega) - rK\,(i\omega), \text{ and } \Delta'\,(i\omega^*) = \left[\frac{d\Delta\,(i\omega)}{d\,(i\omega)}\right]_{\omega=\omega^*} .$$

Dividing both sides of (5.32) by $K(i\omega^*)$, we obtain:

$$I\,(i\omega^*) - R\,(A) = -i\,\frac{\Delta'\,(i\omega^*)}{K\,(i\omega^*)}\,\delta\omega .$$

We now note that

$$\Delta'\,(i\omega^*) = \left[\frac{d\,[D\,(i\omega) - rK\,(i\omega)]}{di\omega}\right]_{\omega=\omega^*} =$$

$$= \left[\frac{d}{di\omega}\;\frac{D\,(i\omega) - rK\,(i\omega)}{K(i\omega)}\;K\,(i\omega)\right]_{\omega=\omega^*} .$$

Differentiating the product and taking into account that

$$\Delta(i\omega^*) = D(i\omega^*) - rK(i\omega^*) = 0 ,$$

we obtain

$$\Delta'\,(i\omega^*) = K\,(i\omega^*)\left[\frac{d}{di\omega}\;\frac{D\,(i\omega) - rK\,(i\omega)}{K\,(i\omega)}\right]_{\omega=\omega^*} =$$

$$= K\,(i\omega^*)\left[\frac{dI\,(i\omega)}{di\omega}\right]_{\omega=\omega^*} .$$

Hence,

$$-\,i\,\frac{\Delta'\,(i\omega^*)}{K\,(i\omega^*)} = -\left[\frac{dI\,(i\omega)}{d\omega}\right]_{\omega=\omega^*} = -\,\Gamma\,(i\omega^*)$$

and equation (5.32) reduces to the form

$$\boxed{I(i\omega^*) = R(A) - \Gamma(i\omega^*)\,\delta\omega.} \tag{5.33}$$

It is now easy to compare the results obtained for the same problem from the method based on the auto-resonance hypothesis with those obtained from that based on the filter hypothesis. In the latter case the amplitude and frequency are found separately from the intersection of the hodographs $I(i\omega)$ and $R(A)$ (Fig. 196). In the former case, as a result of equation (5.33) it is necessary to add to this construction the tangent to the hodograph $I(i\omega)$ at the point $\omega = \omega^*$ corresponding to the propagated frequency, i.e. at the point where this hodograph first intersects the negative real axis (Fig. 197). The amplitude of the steady state is determined by the value of the hodograph at the point where this tangent intersects it, and the correction of the frequency is equal to

$$\delta\omega = \frac{l}{|I'(i\omega^*)|},$$

where l is the segment AB shown in Fig. 197.

The sign of $\delta\omega$ is positive if the direction of the vector $I'(i\omega^*)$ coincides with that of AB, and is negative if their directions are opposed.

From the comparison of the constructions shown in Figs. 196 and 197 it follows that the values of the frequency and amplitude of the auto-oscillatory states determined by the two methods will coincide exactly only when $R(A)$ is a real function of A.* In such cases the application of small parameter methods to systems not containing such small parameters but satisfying the filter hypothesis therefore leads to correct results. The coincidence or non-coincidence with the results of exact methods depends only on how exactly the filter hypothesis is satisfied, and does not depend at all on the magnitude of the small parameter.

It is quite a different matter when $R(A)$ is a complex function of A.** The amplitudes and frequencies found by starting from the

* i.e. when $f(x)$ is single-valued (see p. 354).

** i.e. when $f(x)$ is not single-valued (see p. 354).

auto-resonance hypothesis differ essentially in such cases from those found from the methods based on the filter hypothesis. This difference can be qualitative as well as quantitative.

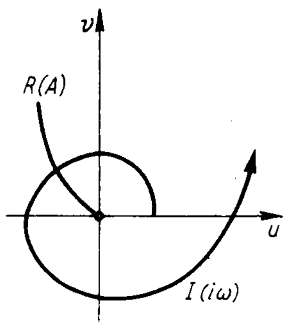


FIG. 196

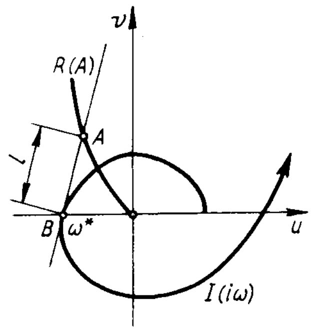


FIG. 197

Thus, for example, in the case given in Fig. 198, the application of auto-resonance hypothesis methods establishes that there are no

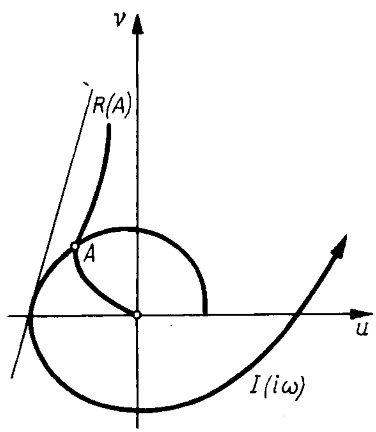


FIG. 198

periodic solutions, whereas the reasoning starting from the filter hypothesis indicates that a periodic state exists (point A in Fig. 198).

These discrepancies do not appear as a consequence of ignoring the harmonics, since they are also ignored in the filter methods. They appear as a result of ignoring terms whose order of smallness exceeds that of the small parameter. The smaller the small parameter, the less this discrepancy will be.

Thus, when $R(A)$ is a complex function of A, the application of methods based on the auto-resonance hypothesis to systems in which the harmonics are ignored on the basis of the filter hypothesis can be the cause of considerable errors.

On the other hand, the application of methods based on the filter hypothesis to systems in which the conditions of the auto-resonance hypothesis are satisfied, not only does not lead to contradictory results, but also gives a more accurate solution.

If, finally, neither the conditions of the auto-resonance nor of the filter hypothesis are satisfied in the system, i.e. if it is not possible to ignore the harmonics at the output of the linear part of the loop, then methods based on either of these hypotheses can produce misleading results.

7. The "Slight" Stability of Periodic Solutions Approximately Found

In the previous sections we have described simple methods of finding the periodic solutions in a number of cases. Only stable periodic solutions, however, correspond to auto-oscillatory states and, consequently, the problem of determining the auto-oscillatory states remains unsolved until the stable states are distinguished from the unstable ones. Forced oscillations in non-linear systems can also be stable or unstable.

Let us return to the system

$$D\left(\frac{d}{dt}\right)x = K\left(\frac{d}{dt}\right)[y + S\sin\omega t], \qquad (5.34)$$

$$y = f(x)$$

or

$$D\left(\frac{d}{dt}\right)x = K\left(\frac{d}{dt}\right)[f(x) + S\sin\omega t] \qquad (5.35)$$

and suppose that a periodic solution $x = f_T(t)$ with period T has been found, such that

$$D\left(\frac{d}{dt}\right)f_T(t) = K\left(\frac{d}{dt}\right)\{f[f_T(t)] + S\sin\omega t\}. \qquad (5.36)$$

To study the stability of these periodic solutions, we will look for x in the form

$$x = f_T(t) + \Delta x.$$

Then

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)[f_T(t) + \Delta x] = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\{f[f_T(t) + \Delta x]\, S \sin \omega t\}.$$

We now expand* the function $f[(f_T(t) + \Delta x]$ in a series of ascending powers of Δx:

$$f[f_T(t) + \Delta x] = f[f_T(t)] + \left[\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right]_{x=f_T(t)} \Delta x + \ldots$$

Being interested only in the "slight" stability of the given periodic state, we shall assume that Δx is so small that all the terms containing it to the second degree and higher can be ignored. Then

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_T(t) + D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\Delta x =$$
$$= K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\left\{f[f_T(t)] + \left[\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right]_{x=f_T(t)} \Delta x + S \sin \omega t\right\}.$$

Taking (5.36) into account, we obtain an equation for determining Δx

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \Delta x = K\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\left[\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right]_{x=f_T(t)} \Delta x\,, \qquad (5.37)$$

where $\left[\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right]_{x=f_T(t)}$ is a given periodic function of time.

Since this equation does not contain S it answers the question of the stability of the periodic state both in the auto-oscillatory case and in the case of forced oscillations.

The given periodic solution is "slightly" stable if the position of equilibrium $\Delta x = 0$ is stable in the system defined by equation (5.37). Thus, the problem of the stability of the periodic solution of a non-

* Here it is assumed that the function $f(x)$ allows such an expansion.

linear system reduces to the problem of determining the stability of the position of equilibrium in the linear system. However, the equation of this linear system contains periodic coefficients, and this is the source of all the difficulties connected with the investigation of stability of periodic states.

We have, so far, succeeded in surmounting these difficulties only when the periodic state being nearly harmonic is stipulated by auto-resonance, and not by a filter.

Bearing in mind the auto-resonance hypothesis, we shall consider further the case when an external periodic action is applied to the system, since the auto-oscillatory case can be obtained from it for $S = 0$.

Let us return to equation (5.34) and, considering small deviations from the periodic state, we shall look for the oscillatory process in the form

$$x = (A_0 + \Delta A) \sin [\omega t - (\gamma_0 + \Delta \gamma)]. \tag{5.38}$$

We put (5.38) in (5.34):

$$\begin{aligned} &D(P) \left[(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta \gamma)\right] = \\ &= K(P) \left\{f[(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma)] + S \sin \omega t\right\}. \end{aligned} \tag{5.39}$$

Here $P = \dfrac{\mathrm{d}}{\mathrm{d}t}$ is the differentiation operator with respect to t.*

Starting from the presence of auto-resonance we can average the function $f[(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma)]$ over the period $\dfrac{2\pi}{\omega}$, approximately representing it thus:**

$$\begin{aligned} &f[(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma)] \approx \\ &\approx (A_0 + \Delta A)\, g(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma) + \\ &+ (A_0 + \Delta A)\, b(A_0 + \Delta A) \cos (\omega t - \gamma_0 - \Delta\gamma)\,, \end{aligned} \tag{5.40}$$

* We must distinguish between $P = \dfrac{\mathrm{d}}{\mathrm{d}t}$ and the complex variable used above in the expressions for the Laplace transform.

** It is assumed that the expansion of $f[A \sin (\omega t - \gamma)]$ does not contain a free term. If there is such a term, and $D(P)$ does not contain P as a factor, the calculations are more complicated.

where $b(A_0 + \Delta A)$ and $g(A_0 + \Delta A)$ are the coefficients of the terms containing $\sin \omega t$ and $\cos \omega t$ in the Fourier expansion of the function

$$f[(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma)].$$

Hence (5.39) reduces to

$$\begin{aligned} &D(P) [(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma)] = \\ &= K(P) [(A_0 + \Delta A) g(A_0 + \Delta A) \sin (\omega t - \gamma_0 - \Delta\gamma) + \\ &+ (A_0 + \Delta A) b(A_0 + \Delta A) \cos (\omega t - \gamma_0 - \Delta\gamma) + S \sin \omega t]. \end{aligned} \tag{5.41}$$

By the usual method of transferring to complex quantities, we obtain:

$$\begin{aligned} D(P)(A_0 + \Delta A) e^{i\omega t} e^{-i(\gamma_0 + \Delta\gamma)} = \\ = K(P)[R(A_0 + \Delta A)(A_0 + \Delta A) e^{i\omega t} e^{-i(\gamma_0 + \Delta\gamma)} + Se^{i\omega t}]. \end{aligned} \tag{5.42}$$

Equation (5.41) is obtained by equating the imaginary parts in (5.42).

Using the easily verified identity

$$S(P)[e^{i\omega t} \varphi(t)] = e^{i\omega t} S(P + i\omega) \varphi(t),$$

where $S(P)$ is any polynomial in P, we transform equation (5.42) to the form

$$\begin{aligned} D(P + i\omega)(A_0 + \Delta A) e^{-i(\gamma_0 + \Delta\gamma)} = \\ = K(P + i\omega)[R(A_0 + \Delta A)(A_0 + \Delta A) e^{-i(\gamma_0 + \Delta\gamma)} + S]. \end{aligned} \tag{5.43}$$

If we assume in (5.43) that $\Delta A = 0$ and $\Delta\gamma = 0$, then we obtain equation (5.19) which served for the determination of the steady oscillations. We are now interested in small deflections from these steady oscillations, so we expand the non-linear functions contained in (5.43) in series in ΔA and $\Delta\gamma$ and ignore the non-linear terms of these series:

$$\begin{aligned} &D(P + i\omega)[A_0 e^{-i\gamma_0} + e^{-i\gamma_0} \Delta A - iA_0 e^{-i\gamma_0} \Delta\gamma] = \\ &= K(P + i\omega)\Bigg[R(A) A_0 e^{-i\gamma_0} + \left[\frac{dR(A)}{dA}\right]_{A=A_0} A_0 e^{-i\gamma_0} \Delta A + \\ &\quad + R(A_0) e^{-i\gamma_0} \Delta A - iR(A_0) A_0 e^{-i\gamma_0} \Delta\gamma + S\Bigg]. \end{aligned} \tag{5.44}$$

Taking the steady state equation

$$D(i\omega)\,A_0 e^{-i\gamma_0} = K(i\omega)\,[R(A_0)\,e^{-i\gamma_0} + S].$$

into account, we reduce equation (5.44) to the form

$$\left[D(P+i\omega) - K(P+i\omega)\,R(A_0) - K(P+i\omega)\left[\frac{\mathrm{d}R(A)}{\mathrm{d}A}\right]_{A=A0} A_0\right]\Delta A -$$

$$- iA_0\,[D(P+i\omega) - K(P+i\omega)\,R(A_0)]\,\Delta\gamma = 0 \qquad (5.45)$$

or

$$- iA_0 M_1 \Delta\gamma + M_2 \Delta A = 0\,, \qquad (5.46)$$

where

$$M_1 = D(P+i\omega) - K(P+i\omega)\,R(A_0)\,,$$

$$M_2 = D(P+i\omega) - K(P+i\omega)\,R(A_0) -$$

$$- K(P+i\omega)\left[\frac{\mathrm{d}R(A)}{\mathrm{d}A}\right]_{A=A_0} A_0. \qquad (5.47)$$

Equation (5.46) is an equation with complex coefficients, which is linear with respect to ΔA and $\Delta\gamma$. Let

$$M_1 = U_1 + iV_1,\ M_2 = U_2 + iV_2.$$

Separating real and imaginary parts in (5.46) we obtain:

$$V_2 \Delta A - A_0 U_1 \Delta\gamma = 0,\ U_2 \Delta A + A_0 V_1 \Delta\gamma = 0\,, \qquad (5.48)$$

where U_1, V_1, U_2, V_2 are polynomials in P with real coefficients.

The characteristic equation of the linear system (5.48) will be

$$\begin{vmatrix} V_2 & - A_0 U_1 \\ U_2 & + A_0 V_1 \end{vmatrix} = A_0 (U_1 U_2 + V_1 V_2)\,. \qquad (5.49)$$

Hence, the equation

$$\mathrm{Re}\,\overline{M}_1 M_2 = 0\,, \qquad (5.50)$$

is the characteristic equation which, by the distribution of its roots relative to the imaginary axis, roughly answers the question of the

stability of the given periodic state, where M_1 is a complex conjugate of M_1, and the multiplication in (5.50) is taken to mean the usual multiplication of complex numbers.

Expanding $D(P + i\,\omega)$ and $K(P + i\,\omega)$ in series in P and putting these series in expressions (5.47) we reduce these to the form:

$$M_1 = \sum_{r=0}^{r=n} \frac{1}{r!} \frac{\partial^r L}{\partial (i\omega)^r} P^r,$$

$$M_2 = \sum_{r=0}^{r=n} \frac{1}{r!} \frac{\partial^r L}{\partial (i\omega)^r} P^r + A_0 \sum_{r=0}^{r=n} \frac{1}{r!} \frac{\partial^r}{\partial (i\omega)^r} \left[\frac{\partial L}{\partial A}\right]_{A=A_0} P^r, \tag{5.51}$$

where

$$L = D(i\omega) - K(i\omega)\, R(A_0). \tag{5.52}$$

If now we put (5.51) in (5.50), the characteristic equation reduces to the form

$$b_0 P^{2n} + b_1 P^{2n-1} + \ldots + b_{2n-1} P + b_{2n} = 0\,, \tag{5.53}$$

where

$$b_0 = \frac{1}{(n!)^2} \operatorname{Re}\left[\frac{\overline{\partial^n L}}{\partial (i\omega)^n} \frac{\partial^n}{\partial (i\omega)^n} \left(L + A_0 \frac{\partial L}{\partial A_0}\right)\right],$$

$$b_1 = \frac{1}{n!\,(n-1)!} \operatorname{Re}\left\{\frac{\overline{\partial^{n-1} L}}{\partial (i\omega)^{n-1}} \left[\frac{\partial^n L}{\partial (i\omega)^n} + A_0 \frac{\partial^n}{\partial (i\omega)^n} \left(\frac{\partial L}{\partial A_0}\right)\right] + \right.$$

$$\left. + \frac{\overline{\partial^n L}}{\partial (i\omega)^n} \left[\frac{\partial^{n-1} L}{\partial (i\omega)^{n-1}} + A_0 \frac{\partial^{n-1}}{\partial (i\omega)^{n-1}} \left(\frac{\partial L}{\partial A_0}\right)\right]\right\},$$

. . . .

$$b_{2n-1} = \operatorname{Re}\left[A_0 \bar{L} \frac{\partial}{\partial (i\omega)} \left(\frac{\partial L}{\partial A_0}\right) + \frac{\overline{\partial L}}{\partial (i\omega)} \left(L + A_0 \frac{\partial L}{\partial A_0}\right) + \bar{L} \frac{\partial L}{\partial (i\omega)}\right],$$

$$b_{2n} = \operatorname{Re}\left|\bar{L}\left(L + A_0 \frac{\partial L}{\partial A_0}\right]\right..$$

The considered approximate solution is "slightly" stable if all the roots of equation (5.53) lie to the left of the imaginary axis.

In an auto-oscillatory state, when $S = 0$, from the steady state equation $L = 0$ and hence $b_{2n} = 0$.

In this case the characteristic equation has a zero root. But from the Andronov—Vitt theorem, this zero root can be rejected, and the characteristic equation reduces to the form

$$b_0 P^{2n-1} + b_1 P^{2n-2} + \ldots + b_{2n-2} P + b_{2n-1} = 0\,, \qquad (5.55)$$

where the coefficients b_i have the values given above with $L = 0$.

In the derivation of the characteristic equations (5.53) and (5.55) we used the auto-resonance hypothesis to justify replacing functions of the form $f[A(t) \sin (\omega t + \gamma(t))]$ by their fundamental components over the period $\frac{2\pi}{\omega}$, but in doing this we did not ignore small quantities in the obtained characteristic equation.

We note now that in the case of auto-resonance $\Delta A(t)$ and $\Delta\gamma(t)$ are slowly changing functions and their derivatives can be considered as being of the second order of smallness.

If the given system is close to a linear system in which a pair of roots of the characteristic equation lies on the imaginary axis and the others to the left of it, we can ignore the higher terms of the characteristic equation (5.53), since all its roots will also lie to the left of the imaginary axis, with the exception of a pair of complex roots whose real part will be determined by the lowest terms of the characteristic equation. This follows directly from the theorem of the continuous dependence of the roots of the characteristic equation on its coefficients and from the fact that the characteristic equation (5.53) continuously tends to the characteristic equation of the linear system when the characteristic $f(x)$ tends to a straight line.

Thus, in the auto-oscillatory case an estimate of the stability can be made from the sign only of the coefficient b_{2n-1} and b_{2n} (in the case of forced oscillations).

When the filter hypothesis was used it was not assumed that the system was nearly linear. Hence, in this case an estimate of the stability of the periodic state solely based on the sign of the coefficient b_{2n-1} or of b_{2n-1} and b_{2n} is not justified. Moreover, in this case equation (5.53) does not solve the question of stability, since the replacement of the non-periodic function by its fundamental over a period cannot now be legitimately justified: $\overline{A}(t)$ and $\gamma(t)$ do not change slowly and in the oscillatory process any frequencies less than ω_c are possible.

Often, when investigating auto-oscillations by the filter method, to determine the stability of the periodic solutions we use the following simple criterion: *a periodic solution, found from the intersection of the hodographs* $I(i\omega)$ *and* $R(A)$, *is stable if the point* $(A_0 + \Delta A)$ *on* $R(A)$ *lies inside the hodograph* $I(i\omega)$ *when* $\Delta A < 0$ *and outside when* $\Delta A < 0$ *(Figs. 199 and 200).* Because of the fact that the increments ΔA are small, this reduces to the condition that the vector tangent to the hodograph $R(A)$ at the given point A_0 is directed into the region bounded by the hodograph $I(i\omega)$. This is equivalent to the requirement that the vector tangent $\left[\frac{\mathrm{d}R(A)}{\mathrm{d}A}\right]_{A=A_0} = T$ shall form an angle less than $\frac{1}{2}\pi$ with the normal to the curve $I(i\omega)$ directed into the given region at the given point of intersection of the two hodographs, i.e. with the vector $i\,\frac{\mathrm{d}I(i\omega)}{\mathrm{d}\omega} = N$. Analytically this condition can be expressed by the use of the scalar product of the two vectors

$$(NT) > 0. \tag{5.56}$$

We return now to the characteristic equation (5.55). Its free term is equal to

$$b_{2n-1} = \mathrm{Re}\left[\frac{\partial \bar{L}}{\partial (i\omega)}\left(2L + A_0\frac{\partial L}{\partial A_0}\right) + A_0\bar{L}\frac{\partial}{\partial (i\omega)}\left(\frac{\partial L}{\partial A_0}\right)\right].$$

We note that

$$\frac{\partial L}{\partial (i\omega)} = \frac{\partial}{\partial (i\omega)}\left(\frac{LK(i\omega)}{K(i\omega)}\right) =$$

$$= \frac{L}{K(i\omega)}\,\frac{\mathrm{d}K(i\omega)}{\mathrm{d}(i\omega)} + K(i\omega)\frac{\partial}{\partial (i\omega)}\left(\frac{L}{K(i\omega)}\right).$$

Now taking it into account that $L = 0$, we obtain:

$$\frac{\partial L}{\partial (i\omega)} = K(i\omega)\frac{\partial}{\partial (i\omega)}\left(\frac{D(i\omega) - R(A_0)K(i\omega)}{K(i\omega)}\right) =$$

$$= K(i\omega)\frac{\mathrm{d}}{\mathrm{d}(i\omega)}\left(\frac{D(\omega)}{K(i\omega)}\right) = K(i\omega)\,\frac{\mathrm{d}I(i\omega)}{\mathrm{d}(i\omega)}$$

and correspondingly

$$\frac{\overline{\mathrm{d}L}}{\mathrm{d}\,(i\omega)} = \overline{K\,(i\omega)}\,\frac{\overline{\mathrm{d}I\,(i\omega)}}{\mathrm{d}\,(i\omega)}$$

Moreover, using (5.52), we find:

$$A_0\frac{\partial L}{\partial A_0} = -\,A_0\,K\,(i\omega)\,\frac{\mathrm{d}R\,(A_0)}{\mathrm{d}A_0}$$

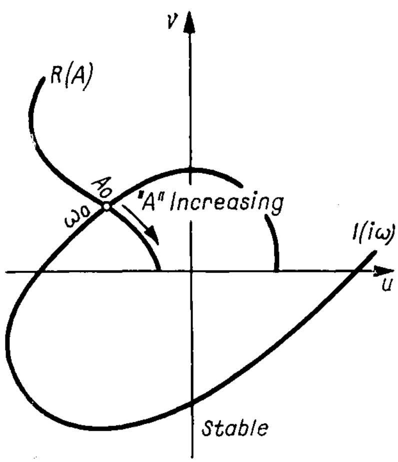


FIG. 199

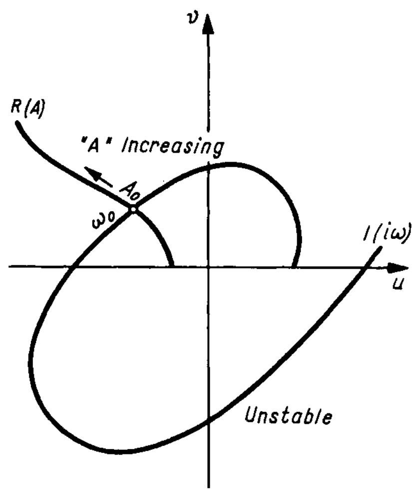


FIG. 200

Hence, finally,

$$b_{2n-1} = -\,A_0\,\overline{K\,(i\omega)}\,K\,(i\omega)\,\mathrm{Re}\left(\frac{\overline{\mathrm{d}I\,(i\omega)}}{\mathrm{d}\,(i\omega)}\;\frac{\mathrm{d}R\,(A_0)}{\mathrm{d}A_0}\right) =$$

$$= -\,A_0\,|K\,(i\omega)|^2\,\mathrm{Re}\left(\frac{\overline{\mathrm{d}I\,(i\omega)}}{\mathrm{d}\,(i\omega)}\;\frac{\mathrm{d}R\,(A_0)}{\mathrm{d}A_0}\right) =$$

$$= A_0\,|K\,(i\omega)|^2\,\mathrm{Re}\left(i\,\frac{\mathrm{d}I\,(i\omega)}{\mathrm{d}\omega}\;\frac{\mathrm{d}R\,(A_0)}{\mathrm{d}A_0}\right).$$

Putting $i\dfrac{\mathrm{d}I(i\omega)}{\mathrm{d}\omega}=N$, $\dfrac{\mathrm{d}R(A_0)}{\mathrm{d}A_0}=T$, we can rewrite this expression in the form of a scalar product:

$$b_{2n-1}=A_0\,|K(i\omega)|^2\,(NT). \tag{5.57}$$

A comparison of (5.56) and (5.57) shows that the frequently used conditions for stability shown in italics above corresponds to the free term of the characteristic equation (5.55) being positive and hence it is suitable only for systems satisfying the conditions of the auto-resonance hypothesis, and does not answer the question of the stability for systems in which the filter hypothesis is realized.

In system in which a filter (and not auto-resonance) ensures that the periodic solutions are nearly harmonic, we cannot use the given simple criterion for stability. Its application in this case will often lead to mistaken conclusions about the stability. Criteria for the stability of periodic states which are based on the filter hypothesis have not yet been found.

B. THE EXACT DETERMINATION OF PERIODIC STATES WHEN THE NON-LINEAR ELEMENT HAS A PIECEWISE-LINEAR CHARACTERISTIC

8. General Introduction to Piecewise-Linear Systems and to Exact Methods for Determining their Periodic Solutions

We return now to the initial system of equations (5.1) and to be specific we put $k=1$.

We shall assume henceforth that the function $f(x_1)$ is made up of straight line segments. Common examples of characteristics of this sort are shown in Fig. 201.

In this part we shall no longer assume that the system satisfies the conditions of either the auto-resonance hypothesis or the filter hypothesis, and therefore we shall take into account all the harmonics given rise to by the non-linear element.

In the system (5.1) we replace $f(x_1)$ by y and obtain an equation containing only x_1 and y, i.e. we eliminate all the other unknowns $x_2, x_3, \ldots, x_n$.

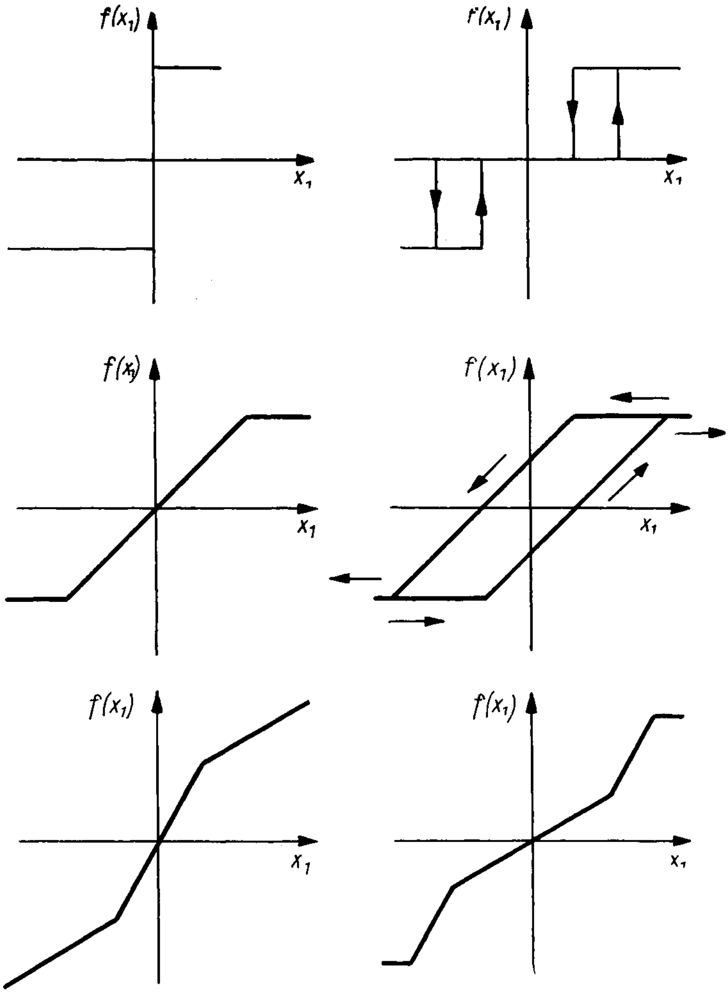


FIG. 201

As a result of this elimination we obtain the equation

$$D(P)\,x_1 = K(P)\,y\,, \tag{5.58}$$

where, as in the previous section, $P = \dfrac{\mathrm{d}}{\mathrm{d}t}$ is the differentiation operator with respect to time.

Equation (5.58) will be called the *derived equation*.

It can be obtained, for example, by transition to a Laplace transform

$$D(p)\,L[x_1] = K(p)\,L[y]\,,$$

where $L[x_1]$ and $L[y]$ are Laplace transforms of the functions x_1 and y respectively. Equation (5.58) is obtained from the last expression by replacing $L[x_1]$ and $L[y]$ by x_1 and y and the complex variable p by the differentiation operator P. But such a substitution can only be made if the functions x_1 and y are sufficiently smooth. In fact, obtaining (5.58) from (5.1) requires the differentiation of these functions (in the Laplace transforms after their multiplication by p), and if they are not smooth, i.e. if in the process of transition from (5.1) to (5.58) we have to differentiate discontinuous functions, then such a transition (or such a substitution of p by P) requires caution.

In this case, there is no meaning even in writing the defined equations in the form (5.58): it is not clear without supplementary clarification what we are to understand by the function $K(P)\,f(x_1)$ if $f(x_1)$ has discontinuities and breaks.

We shall agree to say that the symbol $PF(t)$ denotes the *ordinary derivative* of the function $F(t)$ with respect to t, which exists everywhere apart from points $t = t_q$, for which the function $F(t)$ has discontinuities; at the points for which $t = t_q$ the derivative does not exist and has a meaning only if we speak of the derivative from the right (for $t = t_q + \varepsilon$, $\varepsilon \to 0$, or simply for $t = t_q + 0$) and of a derivative from the left (for $t = t_q - 0$). In Fig. 202 examples of the functions $F(t)$ and their ordinary derivatives $PF(t)$ are given.

We now introduce the Dirac δ-function, which we define in the following way: if $h(t)$ is an arbitrary, sufficiently smooth function,

$$\int_{-\infty}^{+\infty} h(t)\,\delta(t)\,\mathrm{d}t = h(0), \tag{5.59}$$

we can represent the function $\delta(t)$ as the limit of the function shown in Fig. 203 when $a \to 0$ and $b \to \infty$ in such a way that their product $ab = 1$.

We now introduce the operation of the *generalized derivative* of the function $F(t)$, which we denote by $P^*F(t)$ and define thus:

$$P^*F(t) = PF(t) + \sum_q \zeta_{0q}\,\delta(t - t_q), \tag{5.60}$$

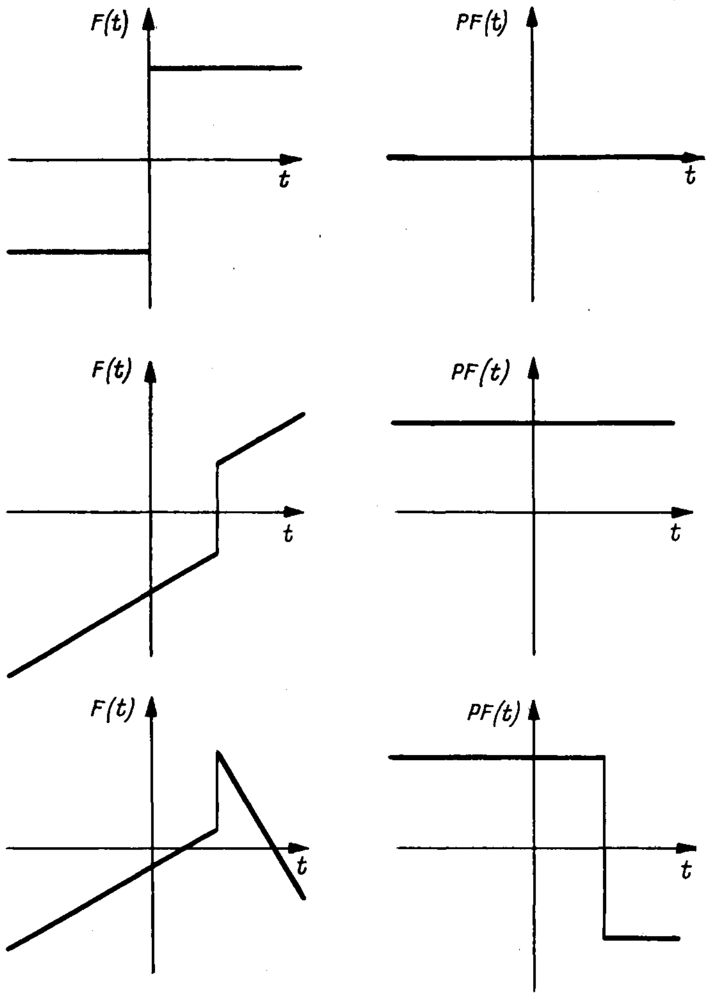


FIG. 202

where the t_q are values of t at which $F(t)$ has discontinuities, and ζ_{0q} is the magnitude of this discontinuity, i. e.

$$\zeta_{0q} = F(t_q + 0) - F(t_q - 0). \tag{5.61}$$

The summation in (5.60) is taken over all the discontinuities.

If the functions $F(t)$ are continuous, then the operations $PF(t)$ and $P^* F(t)$ coincide, since all $\zeta_{0q} = 0$.

Repeatedly applying the operation P^*, we obtain formulae defining the second and higher generalized derivatives:

$$P^* F(t) = PF(t) + \sum_q \zeta_{0q} \delta(t - t_q),$$

$$P^{*2} F(t) = P^2 F(t) + \sum_q [\zeta_{0q} \delta'(t - t_q) + \zeta_{1q} \delta(t - t_q)],$$

$$\dots\dots\dots\dots\dots\dots \qquad (5.62)$$

$$P^{*n} F(t) = P^n F(t) +$$

$$+ \sum_q [\zeta_{0q} \delta^{n-1}(t - t_q) + \zeta_{1q} \delta^{n-2}(t - t_q) + \dots + \zeta_{n-1} \delta(t - t_q)].$$

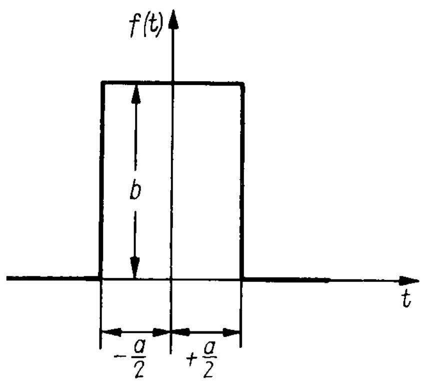


FIG. 203

Here $\zeta_{0q}, \zeta_{1q}, \zeta_{2q}, \dots, \zeta_{nq}$ are the discontinuities of the functions $F(t), PF(t), P^2 F(t), \dots, P^n F(t)$ at the times $t = t_q$ when a discontinuity occurs for any one of these functions. The summation in formula (5.62) is taken over all the points t_q. The functions $\delta', \delta'', \dots, \delta^{n-2}, \delta^{n-1}$ denote the first, second and higher derivatives of the Dirac δ-function. Henceforth all these functions will be ignored and we shall not meet them again in this book. We shall not therefore explain more precisely here how these derivatives are determined.

If we now return to the initial system (5.1) and once again eliminate all the unknown functions apart from x_1, not by ordinary differentiation, but by using the generalized derivative, then as a result we obtain the same derived equation (5.58) except that now the

operator P will be replaced everywhere by P^*:

$$\boxed{\begin{aligned} D(P^*)\,x_1 &= K(P^*)\,y, \\ y &= f(x_1). \end{aligned}} \tag{5.63}$$

We call this equation the *generalized derived equation.*

And now, despite the fact that the function is not smooth, the equation $D(P^*)\,x_1 = K(P^*)\,f(x_1)$ has a clear meaning. Put*

$$\left.\begin{aligned} D(P^*) &= a_0 P^{*n} + a_1 P^{*n-1} + \ldots + a_n, \\ K(P^*) &= b_0 P^{*n} + b_1 P^{*n-1} + \ldots + b_n. \end{aligned}\right\} \tag{5.64}$$

In (5.64) we put the values of $P^*, P^{*2}, \ldots$ taken from (5.62). We then equate the terms on the left and right which do not contain the δ-function, and, for each instant t_q separately, the terms containing δ-functions, δ'-functions, δ''-functions, and so on.

As a result instead of the one generalized reduced equation

$$D(P^*)\,x_1 = K(P^*)\,f(x_1)$$

we obtain the equation

$$\boxed{D(P)\,x_1 = K((P)\,f(x_1)} \tag{5.65}$$

together with the supplementary equations for each instant

$$\boxed{\begin{aligned} & a_0\,\zeta_0 = b_0\,\eta_0, \\ & a_0\,\zeta_1 + a_1\,\zeta_0 = b_0\,\eta_1 + b_1\,\eta_0, \\ & \quad \cdots\cdots\cdots\cdots \\ & a_0\,\zeta_{n-1} + \ldots + a_{n-1}\,\zeta_0 = b_0\,\eta_{n-1} + \ldots + b_{n-1}\,\eta_0, \end{aligned}} \tag{5.66}$$

which we call the *saltus conditions.*

Here $\zeta_0, \zeta_1, \ldots, \zeta_{n-1}$, as before are the discontinuities of the functions $x_1, Px_1, \ldots, P^{n-1}x_1$ at the instant $t = t_q$ (the index q being

* Usually the degree m of the polynomial $K(P^*)$ is less than the degree n of the polynomial $D(P^*)$. But this only means that in (5.64) $b_0, b_1, \ldots$ are equal to zero.

omitted) and $\eta_0, \eta_1, \ldots, \eta_{n-1}$ are similarly the discontinuities of the function $f[x_1(t)]$:

$$\eta_0 = f[x_1(t_q + 0)] - f[x_1(t_q - 0)],$$

$$\eta_1 = Pf[x_1(t_q + 0)] - Pf[x_1(t_q - 0)] = \\ = \left[\frac{df(x_1)}{dx_1}\right]^+ \dot{x}_1^+ - \left[\frac{df(x_1)}{dx_1}\right]^- \dot{x}_1^-$$

The indices "+" and "−" above denote that $x_1(t_q - 0)$ and $x_1(t_q + 0)$ respectively have replaced $x_1(t)$, and $t_q + 0$ and $t_q - 0$ have replaced t_q.

From the point of view of determining the motion of the coordinate x_1 the initial system (5.1) is equivalent either to the single generalized derived equation (5.63), containing a generalized derivative, or to the derived equation (5.65), containing an ordinary derivative, together with the saltus conditions (5.66) for all times $t = t_q$ at which there are discontinuous in any one of the functions

$$x_1(t),\ Px_1(t),\ \ldots,\ P^{n-1}x_1(t),$$

$$f[x_1(t)],\ Pf[x_1(t)],\ \ldots,\ P^{n-1}f[x_1(t)].$$

We denote by $t_1, t_2, t_3, \ldots$ the moments t_q. For $t_1 < t < t_2$, $t_2 < t < t_3$ and so on, the functions x_1 and $f(x_1)$ are smooth.

Let the values of $x_1, Px_1, \ldots, P^{n-1}x_1$ be known for the time $t = t_1 + 0$. Then, taking them as boundary conditions, we can integrate the equation obtained from (5.65) by putting in it the equation of the first stage of the characteristic

$$f(x_1) = S_1 x_1 + g_1,$$

and determine x_1 for the first interval ($t_1 < t < t_2$).

From this integral are found the values of the functions x_1, $Px_1, \ldots, P^{n-1}x_1$ at the time $t = t_2 - 0$. It is not possible to take these as the boundary conditions for the following step, since for $t = t_2$ the function x_1 and (or) its derivatives have discontinuities. But putting these values of $x_1, Px_1, \ldots, P^{n-1}x_1$ and $f(x_1), Pf(x_1), \ldots,$ $P^{n-1}f(x_1)$ calculated from them for the time $t = t_2 - 0$ in the saltus conditions (5.66) we obtain from them n linear relations, determining $x_1, Px_1, \ldots, P^{n-1}x_1$ at the time $t = t_2 + 0$.

Then again, taking these as boundary conditions, we can integrate the derived equation (5.65). Taking y from the equation of the second stage of the characteristic, we can determine in this way the motion during the second step ($t_2 < t < t_3$) and find $x_1, Px_1, \ldots, P^{n-1} x_1$ for $t = t_3 - 0$. Then, repeatedly applying the saltus conditions (5.66) we find n new relations for determining the boundary conditions of the subsequent, third, step, and so on.

Thus, the generalized derived equation (5.63), or the derived equation (5.58), supplemented by the saltus conditions, completely determines the change in x_1 with time for the given boundary conditions.

The operation of step integration which we have described above, supplemented, if need be, by the calculation of the discontinuities from the saltus conditions, is called *approvision**.

Approvision discloses a way of determining exactly the periodic states (without ignoring harmonics). For this it is necessary only that, in carrying it out over a period, the values of $x_1, Px_1, \ldots, P^{n-1} x_1$ at the end of the period T (for $t = T + 0$ if for $t = T$ there is a discontinuity) shall coincide with the boundary conditions (for $t = +0$ if there is a discontinuity for $t = 0$).

Three kinds of difficulty arise in the use of this method.

In order to explain the first of these, let us consider the simplest example of a system having a symmetrical relay characteristic (Fig. 204).

We can picture a periodic state in which the depicted point, as it moves along the characteristic during the period time T successively passes the points 1, 2, 3, 4, 3, 2, 1 and then repeats the cycle. For this the relay is switched twice during the period. But we can also imagine another state, where during one period the points 1, 2, 3, 4, 3, 2, 5, 2, 3, 6, 3, 2, 1, are passed in succession, i.e. the relay is switched four times per period.

* We could integrate the initial system (5.1) directly. Then, during "approvision" it would not be necessary to use the saltus conditions, but, on the other hand, we would have to integrate at each step not one, but a system of equations. The calculations which such a method of "approvosion" would entail would not be simpler than those we have described, while the use of the derived equation is more convenient for control theory problems. The technicalities of approvision, taking the saltus conditions into account, are described in Section 4 of the work: Aizerman, M. A. and Gantmakher, F. R., On the Determination of Periodic States in a Non-linear Dynamic System with Piecewise-Linear Characteristic, *Prikl. mat. i mekh.* XX, No. 5 (1956).

Of course, even more complicated periodic states with a larger number of relay switchings during a period are also possible. Similarly with other piecewise-linear characteristics. Various types of periodic state are possible in the same system differing in the order in which the separate lines constituting the characteristic are passed during each period. But to carry out approvision it is necessary to know beforehand in what order these lines occur, i.e. in what order the various linear equations (5.58) describing the motions in separate steps alternate with one another.

Because of this we can find all the periodic solutions by the approvision method in any system with a piecewise-linear characteristic only by trial and error: it is necessary to fix a definite type of periodic state, i.e. a definite order in which the stages of the characteristic occur over a period, and to determine the periodic motions of this type by the approvision method; then in the same way to find the periodic motion of another type and so successively to try all possible types of periodic state.*

The second difficulty associated with the approvision method for determining the period states consists in the following.

In each step (the first, say) (i.e. for $t_1 < t < t_2$) the change in x_1 is determined by the integral of a linear equation, i.e.

$$x_1 = \sum_{j=1}^{n} C_j \, e^{\lambda_j t},$$

where λ_j are the roots of the characteristic equation for this step Therefore at the beginning of the step

$$x_{1.1} = \sum_{j=1}^{n} C_j \, e^{\lambda_j t_1}, \tag{5.67}$$

and at the end of this step

$$x_{1.2} = \sum_{j=1}^{n} C_j \, e^{\lambda_j t_2}. \tag{5.68}$$

* We note that the existence of periodic states of various types is itself made possible by the presence of harmonics. The approximate methods considered in the previous sections do not allow us to take these differences into account. Therefore, for example, in the case of a non-linear characteritsic and the presence of a filter the harmonic balance method finds only the simplest state with two switchings of the relay per period. Any other types of periodic state $\left(\text{possible only for frequencies } \omega < \frac{\omega_C}{S}\right)$ are not discorvered by this method.

Exactly similar relations can be written out for all the other steps and added to the relations given by the saltus conditions. The values of x at the beginning and end of the steps and of all the C_j enter linearly into these relationships, but the $t_1, t_2, \ldots$ are unknown, and the period T is non-linear. Although the number of equations which can be formed in this way is equal to the number of unknowns, it is, therefore, not possible to find a general solution for these equations. It is possible only by methods of linear algebra to eliminate all the unknowns which enter into the equations linearly and to form a system

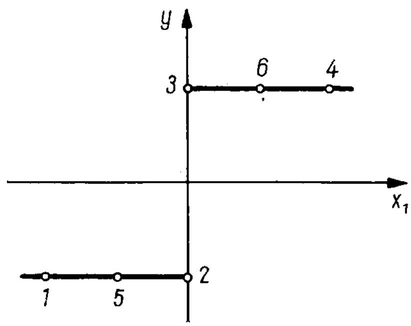


FIG. 204

of non-linear (transcendental) equations with respect to the unknowns T and $t_1, t_2, \ldots$ — the period and its parts corresponding to the separate stages of the characteristic.

This system of equations is called the *period equations*. The approvision method does not enable us to form the period equation which answers the question of the presence or absence of periodic solutions of any type. It only enables us to form the period equations for each of the possible types of periodic solution separately. To solve them (they are usually a complicated system of transcendental equations) it is necessary to use numerical methods, graphical means, or else computing machines.

The third difficulty in applying the approvision method arises because the relations (5.67), (5.68) and, hence, those in the period equations derived from them, involve the roots λ of the characteristic equations. To form the period equations by approvision it is necessary to find the roots of the characteristic equations, of which there are as many as there are stages in the characteristic.

Besides the approvision method there is also another method for determining exactly the periodic states in piecewise-linear systems — they are looked for in the form of complete Fourier series, without neglecting the harmonics :

$$x_1 = \sum_{r=-\infty}^{\infty} a_r e^{ir\omega t}, \qquad y = \sum_{r=-\infty}^{\infty} \beta_r e^{ir\omega t}. \tag{5.69}$$

In this case ω and the two infinite sequences of Fourier coefficients a_r and β_r $(-\infty < r < +\infty)$ have to be determined.

The connexion between a_r and β_r is obtained by substituting the series (5.69) in the generalized derived equation (5.63) and in the equation of the characteristic $y = f(x_1)$.

From the equation of the characteristic we obtain :

$$\sum_{r=-\infty}^{\infty} \beta_r e^{ir\omega t} = f\left(\sum_{r=-\infty}^{\infty} a_r e^{ir\omega t}\right). \tag{5.70}$$

It is then necessary to find the Fourier expansion of the function of the Fourier series on the right hand side of this equation. In the general case when $f(x_1)$ is not piecewise-linear, but is an arbitrary function, all the β_r are expressed from (5.70) non-linearly in a_r and this method leads only to the formation of an infinite system of non-linear equations in the unknown a_r. In the case when $f(x_1)$ is a piecewise-linear function, however, we can ignore this difficulty, since the whole infinite sequence of unknown Fourier coefficients a_r can be expressed linearly by n parameters. Unfortunately, in using this method we cannot ignore the first two difficulties connected with the approvision method. As in approvision we must carry out the solution of the problem separately for various types of periodic state and the solution leads only to the formation of period equations. But in contrast to the approvision method, the period equations formed in this way do not involve the roots of the characteristic equations, but are expressed directly in terms of the coefficients $a_0, a_1, \ldots, a_n$ and $b_0, b_1, \ldots, b_n$ of the derived equations.

Before describing this method, we make the following preliminary remark. Let $f(t)$ be a periodic function and let its Fourier series be known :

$$f(t) = \sum_{r=-\infty}^{\infty} \gamma_r e^{ir\omega t}. \tag{5.71}$$

Then the Fourier series which is obtained by term-by-term differentiation of the series (5.71), i.e.

$$\sum_{r=-\infty}^{\infty} (ir\omega)\,\gamma_r\, e^{ir\omega t},$$

determines not the function $Pf(t)$ but the function $P^*f(t)$.

Thus, for example, if $f(t)$ is a periodic sequence of rectangular pulses (Fig. 205),

$$f(t) = \frac{2}{\pi} \sum_{r=-\infty}^{\infty} \frac{1}{ir}\, e^{ir\omega t},$$

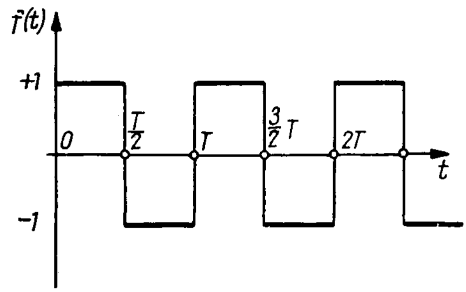


Fig. 205

and the series

$$\frac{2\omega}{\pi} \sum_{r=-\infty}^{\infty} e^{ir\omega t}$$

determines not $Pf(t) \equiv 0$ but $P^*\,f(t)$, i.e. the periodic sequence of the δ-function* (Fig. 206).

From this it follows that the Fourier series of the function

$$L(P^*)\,f(t),$$

where $L(P^*)$ is a polynomial, can be found from the series for $f(t)$ carrying out the operation $L(P^*)$:

$$L(P^*)\,f(t) = \sum_{r=-\infty}^{\infty} L(ir\omega)\,\gamma_r\, e^{ir\omega t}.$$

* To be more precise, the δ-function multiplied by 2, for the magintude of the discontinuity of $f(t)$ is equal to $1 - (-1) = 2$.

Bearing this remark in mind, we proceed to the determination of the period solutions in the form of complete Fourier series (5.69).

In Section 9 this will be done for a very elementary case (the simplest state for a symmetric relay characteristic), in Section 10 first for the simplest periodic state with an arbitrary two-stage characteristic, and then for an arbitrary state for any characteristic consisting of segments parallel to two given straight lines.

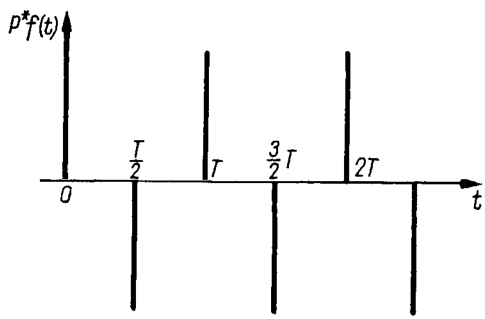


FIG. 206

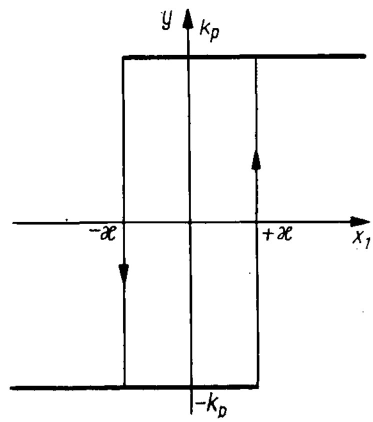


FIG. 207

9. The Simplest Periodic States in a System with a Symmetric Relay

We consider now the system (5.1) or its equivalent generalized derived equation

$$D(P^*)\, x_1 = K(P^*)\, f(x_1)\,. \qquad (5.72)$$

Here $f(x_1)$ is the characteristic of the relay (Fig. 207). In particular cases we can put $\varkappa = 0$ and then the characteristic will have no loop (Fig. 208). The degree m of the polynomial $K(P^*)$ is less than the degree n of the polynomial $D(P^*)$.

We shall look for the most simple periodic state for which the relay switches only at the times which are multiples of $\frac{T}{2}$, where T is the period of the oscillations.

Figure 209 shows the change in $x_1(t)$ and $y(t)$ with time for such a periodic state. In order that the periodic functions $x_1(t)$ and $y(t)$ with a total period T satisfy the equation of the characteristic $y = f(x_1)$ in Fig. 209, they must satisfy the following three conditions :

$$\left.\begin{aligned}
&1^\circ \cdot y(t) = \begin{cases} +k_p \text{ for } 0 < t < \dfrac{T}{2}, \; T < t < \dfrac{3}{2}T \text{ and so on} \\ -k_p \text{ for } \dfrac{T}{2} < t < T, \dfrac{3}{2}T < t < 2T \text{ and so on} \end{cases} \\
&2^\circ \cdot x_1(0) = +\varkappa, \\
&3^\circ \cdot x_1(t) > -\varkappa \text{ for } 0 < t < \frac{T}{2}, \; T < t < \frac{3}{2}T \text{ and so on} \\
&\quad x_1(t) < \varkappa \text{ for } \frac{T}{2} < t < T, \; \frac{3}{2}T < t < 2T \text{ and so on}
\end{aligned}\right\} \tag{5.73}$$

We shall look for the periodic solution in the form of Fourier series

$$\left.\begin{aligned}
x_1 &= \sum_{r=-\infty}^{\infty} \alpha_r e^{ir\omega t}, \\
y &= \sum_{r=-\infty}^{\infty} \beta_r e^{ir\omega t}.
\end{aligned}\right\} \tag{5.74}$$

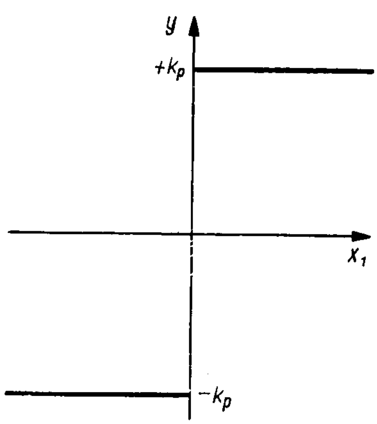


FIG. 208

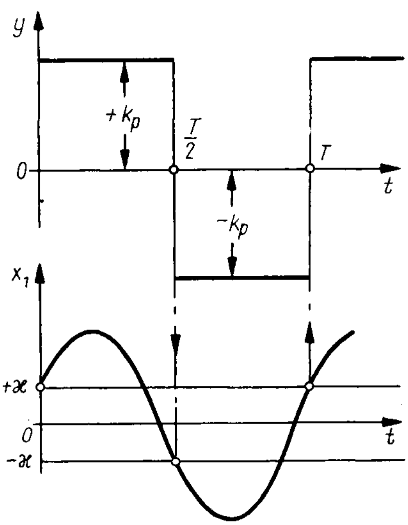


FIG. 209

Putting these series in the derived equation (5.72) and taking the remark made at the end of the previous section concerning the Fourier series of $L(P^*)\,f(t)$ into account, we obtain

$$\sum_{r=-\infty}^{\infty} a_r D\,(i\omega)\, e^{ir\omega t} = \sum_{r=-\infty}^{\infty} \beta_r K\,(ir\omega)\, e^{ir\omega t}.$$

From this we find the relations between all the Fourier coefficients a_r and β_r:

$$a_r = \beta_r \frac{K\,(ir\omega)}{D\,(ir\omega)} = \beta_r W\,(ir\omega), \tag{5.75}$$

where the function $W\,(i\omega) = \dfrac{K\,(i\omega)}{D\,(i\omega)}$ is the frequency characteristic of the linear part of the system.

We now put the series (5.74) in the equation of the characteristic $y = f(x_1)$. In order that it shall be satisfied, the three conditions given above must be satisfied.

The periodic function shown in Fig. 209 has the Fourier series

$$y = \frac{4k_{\mathrm{p}}}{T} \sum_{r=-\infty}^{\infty} \frac{e^{ir\omega t}}{ir\omega}. \tag{5.76}$$

The period solutions found have Fourier coefficients

$$\beta_r = \frac{4k_{\mathrm{p}}}{ir\omega T} \tag{5.77}$$

and

$$a_r = \frac{4k_{\mathrm{p}}}{ir\omega T} \frac{K\,(ir\omega)}{D\,(ir\omega)}. \tag{5.78}$$

All the unknowns a_r are now expressed in terms of one unknown, the period $T\left(\text{we recall that } \omega = \dfrac{2\pi}{T}\right)$.

To find T we note that as yet only condition 1° of (5.73) has been satisfied, and it is still necessary that

$$x_1\,(t) = \sum_{r=-\infty}^{\infty} \cdot\, a_r e^{ir\omega t} = \frac{4k_{\mathrm{p}}}{T} \sum_{r=-\infty}^{\infty} \frac{K\,(ir\omega)}{ir\omega D\,(ir\omega)} e^{ir\omega t} \tag{5.79}$$

shall satisfy the second and third of these conditions.

From the second condition it follows that

$$\varkappa = \frac{4k_p}{T} \sum_{r=-\infty}^{\infty} \frac{K(ir\omega)}{ir\omega D(ir\omega)} \, . \tag{5.80}$$

This is the required period equation.

Let $T = T^*$ be its root. Then

$$x_1 = \frac{4k_p}{T^*} \sum_{r=-\infty}^{\infty} \frac{K(ir\omega^*)}{ir\omega^* D(ir\omega^*)} e^{ir\omega^* t}, \tag{5.81}$$

$$y = \frac{4k_p}{T^*} \sum_{r=-\infty}^{\infty} \frac{e^{ir\omega^* t}}{ir\omega^*}, \tag{5.82}$$

where $\omega^* = \dfrac{2\pi}{T^*}$.

These series satisfy the derived equation and the first two of the three conditions which it was required to fulfil so that the equation of the characteristic should be satisfied. To find whether the third of these conditions is satisfied we have to construct the function (5.81) over the length of one period.

Thus, *the simplest periodic states of a relay are determined by formulae (5.81) and (5.82) where T^* is any root of the period equation (5.80) for which the function (5.81) satisfies condition 3°.*

This last condition gives a solution, and not, as it seems at first glance, only a trivial result. We find examples when the period equation has an infinite number of real roots, all of which are false, since they do not satisfy condition 3°, or only one or two of them satisfy this condition. The roots of the period equation are therefore not the required values of the periods, but are only "pretenders", which include also the required periods which must be found by using condition 3°.

When $\varkappa = 0$, to satisfy condition 3° it is necessary (but not sufficient) that $\left|\dfrac{dx_1}{dt}\right|_{t=0} > 0$.

From (5.81) we obtain:

$$\frac{dx_1}{dt} = \frac{4k_p}{T^*} \sum_{r=-\infty}^{\infty} \frac{K(ir\omega^*)}{D(ir\omega^*)} e^{ir\omega^* t} \, . \tag{5.83}$$

Putting $t = 0$, we find the condition

$$\sum_{r=-\infty}^{\infty} \frac{K(ir\omega^*)}{D(ir^*)} > 0. \tag{5.84}$$

This condition allows the preliminary "rejection" of roots of the period equation for $\varkappa = 0$. Those of them for which the inequality (5.84) is not satisfied must be rejected at once. But all the roots of the

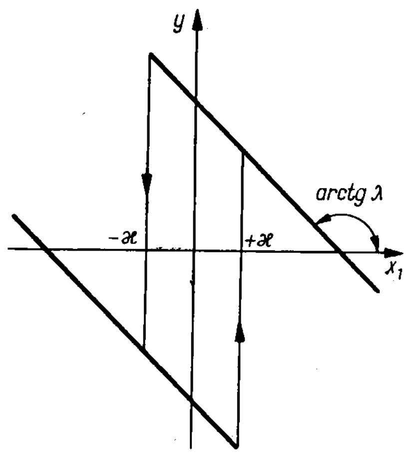


FIG. 210

period equation which satisfy the inequality (5.84) must be checked with respect to condition 3°, as we have described above.

We note in conclusion that the system with the characteristic shown in Fig. 210 (where the slope of the straight lines can be equal to any finite number λ) reduces to the relay system we have considered. For this, in the system of equations

$$D(P^*)\,x_1 = K(P^*)\,y,$$

$$y = f(x_1),$$

we must put

$$y = \lambda\,x_1 + \varphi\,(x_1)\,,$$

where $\varphi\,(x_1)$ is the relay characteristic.

As a result we obtain

$$[D(P^*) - \lambda\, K(P^*)]\, x_1 = K(P^*)\, y_1\,,$$
$$y_1 = \varphi\,(x_1)\,.$$

This is the equation of a relay system since

$$D(P^*) - \lambda\, K(P^*)$$

can be considered as a new polynomial $D(P^*)$.

10. The Determination of the Periodic States in System with a Non-Relay Piecewise-linear Characteristic*

For the sake of simplicity we spoke in Sections 8 and 9 only of the determination of auto-oscillations. Now, coming to the consideration of the characteristics of more general form, we consider at the same time both the determination of auto-oscillations, and that of forced oscillations, whose frequency coincides with that of the external disturbance. Instead of the system of equations (5.1), therefore, we shall now consider the more general system :

$$\dot{x}_j = \sum_{k=1}^{n} a_{jk} x_k + [\lambda_j f(x_1) + F_j(t)\, (j = 1, 2, \ldots, n)\,, \qquad (5.85)$$

where a_{jk} and λ_j are constants (of which several can be equal to zero), $f(x_1)$ is a piecewise-linear function, and the $F_j\,(t)$ are given, sufficiently smooth, periodic functions of time with a total period T (in particular, the $F_j\,(t)$ can be constants either different from or equal to zero).

It is required to determine the periodic solutions of the equations (5.85). In the case when all $F_j\,(t) =$ const. (i.e. in the auto-oscillatory case), we look for all the periodic solutions of the system (5.85). If any one of the $F_j\,(t) \neq$ const., we have to determine only the periodic solutions of the equations (5.85) with period T.

* This section reproduces practically unchanged the work: Aizerman, M. A. and Gantmakher, F. P., The Determination of the Periodic States in Systems with a Piecewise-linear Characteristic, Formed for γ Stages Parallel to Two Given Straight Lines, *Avtomatika i Telemekhanika* Nos. 2 and 3 (1957).

The transformation of the equations

In the system (5.85) we eliminate all the unknowns except x_1 and $f(x_1) = y_1$. As a result we obtain the generalized derived equation

$$\begin{aligned} D(P^*)\, x_1 &= K(P^*)\, y_1 + \Phi(t), \\ y_1 &= f(x_1). \end{aligned} \tag{5.86}$$

The function $\Phi(t)$ is obtained from $F_j(t)$ during the elimination of $x_2, x_3, \ldots, x_n$ by differentiation, multiplication by constants and summation. Therefore $\Phi(t)$ has the same period T; $\Phi(t) \equiv$ const if all $F_j(t) \not\equiv$ const. and, generally speaking $\Phi(t) \not\equiv$ const. if any one of the $F_j(t) \equiv$ const.

We restrict ourselves for the present to a consideration of the characteristic $f(x_1)$ formed by *two given straight lines*. Then the coordinate x_1 is continuous (since $m < n$ and y_1 can have discontinuities, since the transition from one straight line of the characteristic to the other can occur not only at the point of intersection of the lines but also for any given values of x_1 by an instantaneous jump (Fig. 211).

In the special case when both generators of the characteristic are straight lines parallel to the x-axis, we obtain a relay characteristic. We shall look for the simplest periodic solution, for which the transition from one straight line to the other occurs at the beginning of the period ($t = t_0$) at its end ($t = t_2$), and only once during it ($t = t_1$). In other words, when the origin of the period is properly chosen, during one period the depicted point traverses a segment first of one and then of the other line of the characteristic.

We denote by σ_1 and σ_2 the given values of x_1 for which the point crosses over from the first to the second straight line, and from the second to the first respectively*.

Here, we assume in addition that the point crosses from one line to the other at the moment when in its motion along either of the straight lines, the coordinate, x_1 first attains the value σ_1 or σ_2 respectively**.

* In the special case when $\sigma_1 = \sigma_2 = \sigma$, i. e. when the points of intersection of the straight lines coincide, the function y_1 is continuous.

** In future, as was done in the theory of relay systems (Section 9), this supplementary fact is taken into account only at the very end. The periodic solutions are looked for independently of this condition, and then it is checked that the periodic states which are found satisfy it.

If $\Phi(t) = \text{const}$ (in particular, if $\Phi(t) = 0$) then the origin of the period t_0 can be chosen arbitrarily. In future we shall take $t_0 = 0$, and the problem reduces to the determination of the times t_1 and t_2 at which the depicted point is respectively on the first and second branches.

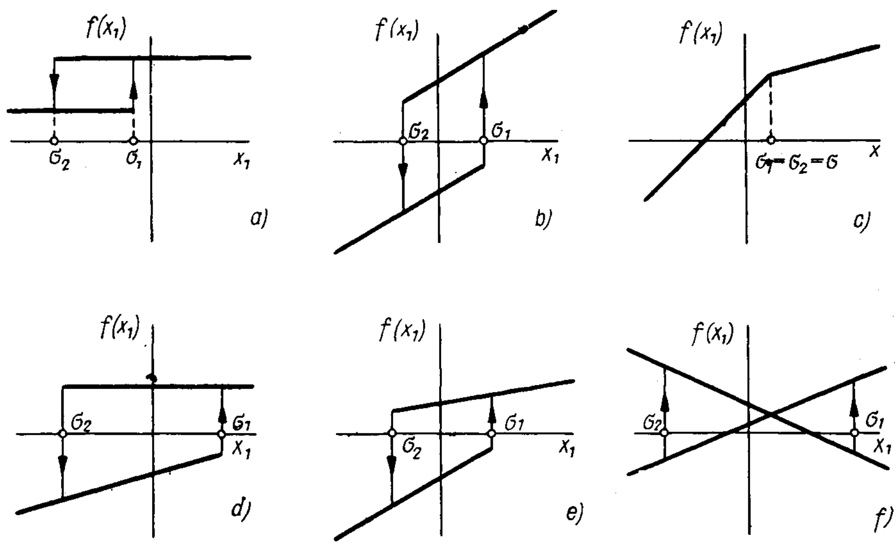


FIG. 211

In the case when $\Phi(t) \neq$ const., the origin of time is determined by the given function $\Phi(t)$ and cannot be selected arbitrarily. The problem reduces again to the determination of two quantities, for example t_1 and t_2 in so far as t_0 is determined from the realtion $T = t_2 - t_0$.

We return to equation (5.86) and make the substitution of the variables

$$\begin{aligned} x_1 &= a\,x + \beta\,y + \varkappa, \\ & \qquad\qquad\qquad (a\delta - \beta\gamma) \neq 0. \\ y_1 &= \gamma\,x + \delta\,y + \lambda, \end{aligned} \tag{5.87}$$

Then instead of (5.86) we have

$$L(P^*)\,x = M(P^*)\,y + \Psi(t), \quad y = F(x), \tag{5.88}$$

where*

$$L(P^*) = aD(P^*) - \gamma K(P^*) = a_0 P^{*n} + a_1 p^{*n-1} + \ldots + a_n,$$
$$M(P^*) = \delta K(P^*) - \beta D(P^*) = b_0 p^{*m} + b_1 p^{*m-1} + \ldots + b_m,$$
$$\Psi(t) = \Phi(t) + K(0)\lambda - D(0)\varkappa.$$

The function $\Psi(t)$ is periodic** of period T. We denote its Fourier coefficients by ε_r, i.e.

$$\Psi(t) = \sum_{r=-\infty}^{\infty}{}' \varepsilon_r e^{ir\omega t}.$$

The function $y = F(x)$ is given explicitly by the relation

$$\gamma x + \delta y + \lambda = f(ax + \beta y + \varkappa).$$

The equations (5.87) determine the linear transformation transforming the points of the x_1, y_1-plane into points of the x, y-plane, and vice versa. If the characteristie consists of two or more sections of straight lines, parallelism is preserved. We select the coefficients a, β, γ, δ, $\varkappa$ and λ of this transformation so that the first straight line of the characteristic is transformed in the x, y-plane into the x-axis, and the second straight line into the y-axes.

To do this we must put (Fig. 212)*

$$\gamma = k_1 a, \qquad \delta = k_2 \beta,$$
$$\varkappa = \frac{h_1 - h_2}{k_2 - k_1}, \qquad \lambda = \frac{k_2 h_1 - k_1 h_2}{k_2 - k_1}.$$

The values of a, β, γ and δ can be selected arbitrarily, provided only that the first two relations in (5.89) are satisfied.

As a result, the equation

$$D(P^*) x_1 = K(P^*) y_1 + \Phi(t)$$

is transformed into

$$L(P^*) x = M(P^*) y + \psi(t),$$

* In contrast to the previous section, here the letters a and b are used to denote the coefficients of the polynomials L and M and not of D and K.

** If $\Phi(t) = 0$ then $\Psi(t) = \text{const} = K(0)\lambda - D(0)\varkappa$.

* These formulae are formed for the case when neither of the straight lines is parallel to the y-axis and when they are not parallel to one another ($k_1 \neq k_2$). In the contrary case the formulae must be modified.

and the characteristic formed from the two non-parallel straight lines transforms into a characteristic formed of segments of the coordinate axes.

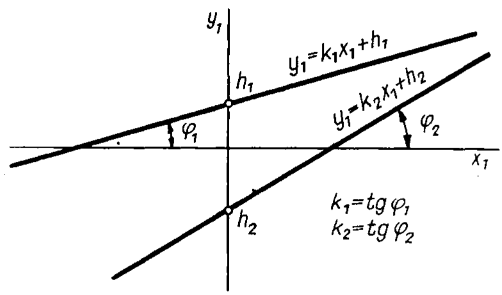


FIG. 212

Earlier, in the (x_1, y_1)-plane, the crossing from the first straight line to the second was made from the point P_1 to the point P_1'. Now, in the x, y-plane, the points Q_1 and Q_1' correspond to these points (Fig. 213).

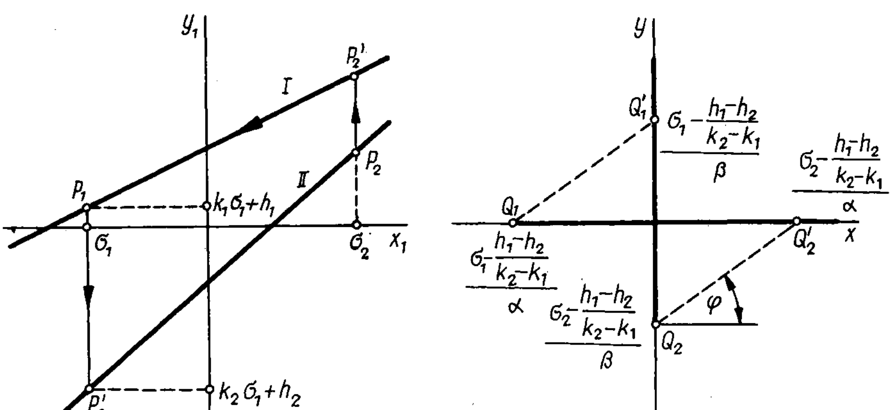


FIG. 213

Similarly, the crossing from the second straight line to the first takes place in the (x, y)-plane from the point Q_2 to the point Q_2'. The coordinates of these points are shown in Fig. 213.

If in the (x_1, y_1)-plane the "saltuses" of a discontinuous state occur along a straight line parallel to the y_1-axis, then in the x, y-plane

saltuses occur along the parallel, but sloping, straight line with a slope $\tan\varphi = \frac{\alpha}{\beta}$.

All this is true only in the case when $k_2 - k_1 \neq 0$, i.e. when the straight lines forming the characteristic are not parallel to one another. We shall not consider here the case when the straight lines are parallel (relay characteristics, or those which reduce to them), since the methods for determining the periodic solutions in relay systems were considered in Section 9.

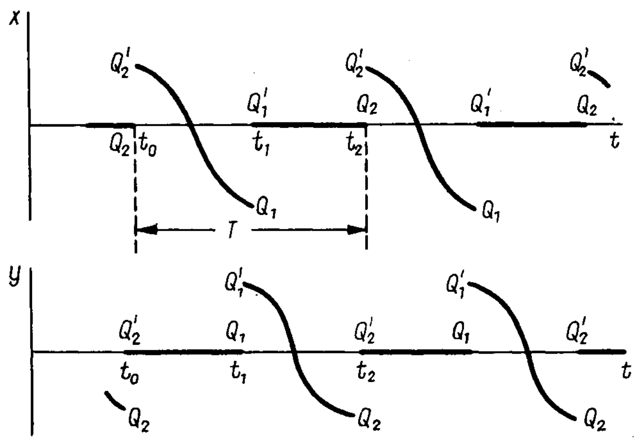


FIG. 214

The formation of the period equation and the determination of the periodic state

The change in the coordinates x and y with time for the simplest periodic state (this term was explained above) for $t_0 < t < t_2$ occurs as in Fig. 214.

We will look for this periodic solution in the form of Fourier series

$$x = \sum_{r=-\infty}^{+\infty} a_r e^{ir\omega t},$$

$$y = \sum_{r=-\infty}^{+\infty} \beta_r e^{ir\omega t}.$$

Putting these series in the equation of the system (5.88), we obtain :

$$L\,(ir\omega)\,a_r = M\,(ir\omega)\,\beta_r + \varepsilon_r.$$

We denote by μ_r^* the Fourier coefficients of the periodic function $M\,(P^*)\,y$, i.e. $\mu_r^* = M(ir\omega)\,\beta_r$. Then

$$a_r = \frac{\mu_r^* + \varepsilon_r}{L\,(ir\omega)}, \qquad \beta_r = \frac{\mu_r^*}{M\,(ir\omega)}. \tag{5.90}$$

Instead of $M(P^*)\,y$ we now consider the function $M(P)y$, where P is the ordinary and not the generalized derivative. For periodic states when y is a function with period T, $M(P)y$ is also a periodic function with the same period. We denote its Fourier coefficients by μ_r.

There exists the following relation between the Fourier coefficients μ_r^* and μ_r in the general case when at the ends of the period there are any number of discontinuities in the function y and its derivatives up to the $(n-1)$th order :

$$\mu_r^* = \mu_r + \frac{1}{T}\left[e^{-ir\omega t_1}\sum_{k=0}^{n-1}\overset{1}{\eta}_k\, m_k\,(ir\omega) + e^{-ir\omega t_2}\sum_{k=0}^{n-1}\overset{2}{\eta}_k\, m_k\,(ir\omega) + \dots\right]. \tag{5.91}$$

Here $\overset{1}{\eta}_k$, $\overset{2}{\eta}_k$ and so on are the magnitudes of the discontinuities of the kth derivative of the function $y(t)$ at the times $t_1, t_2, \dots$,

$$m_k\,(S) = b_{n-k-1} + b_{n-k-2}\,S + \dots + b_0\,S^{n-k-1}$$

and the number of sums in the square brackets is equal to the number of points of discontinuity at the ends of the period. In the case we are considering the number of such points is equal to two (for the times t_1 and t_2) and the terms denoted by dots in the square brackets are absent.

To derive this relationship, we consider the periodic function $\varDelta\,(t) = \sum_{r=-\infty}^{\infty} \delta(t + rT)$ (Fig. 215). Its Fourier expansion is

$$\varDelta\,(t) = \frac{1}{T}\sum_{r=-\infty}^{\infty} e^{ir\omega t}.$$

Supposing that y is a periodic function and that $\overset{1}{\eta}_k$, $\overset{2}{\eta}_k$, ... are the magnitudes of the discontinuities of its kth derivative ($k = 0, 1, \ldots, n-1$) at the ends of the period, we can write the equations

$$y = y,$$

$$P^* y = Py + \sum_q \overset{q}{\eta}_0 \Delta(t - t_q),$$

$$P^{*2} y = P^2 y + \sum_q [\overset{q}{\eta}_0 \Delta'(t - t_q) + \overset{q}{\eta}_1 \Delta(t - t_q)],$$

$$P^{*3} y = P^3 y + \sum_q [\overset{q}{\eta}_0 \Delta''(t - t_q) + \overset{q}{\eta}_1 \Delta'(t - t_q) + \overset{q}{\eta}_2 \Delta(t - t_p)]$$

and so on, up to $P^{*n-1} y$.

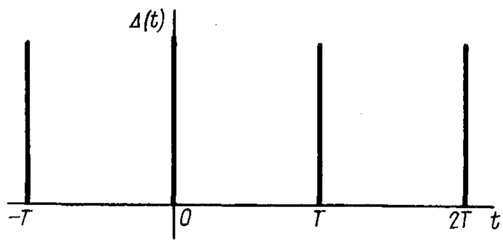


FIG. 215

Multiplying the left and right-hand sides of the first equation by b_n, of the second equation by b_{n-1}, of the third by b_{n-2}, and so on, and adding all the equations, we obtain :

$$M(P^*) y = M(P) y + \sum_q \Delta(t - t_q)(b_{n-1} \overset{q}{\eta}_0 + b_{n-2} \overset{q}{\eta}_1 + \ldots) +$$

$$+ \sum_q \Delta'(t - t_q)(b_{n-2} \overset{q}{\eta}_0 + b_{n-3} \overset{q}{\eta}_1 + \ldots) + \ldots$$

But

$$M(P^*) y = \sum_{r=-\infty}^{\infty} \mu_r^* e^{ir\omega t}, \quad M(P) = y \sum_{r=-\infty}^{\infty} \mu_r e^{ir\omega t},$$

$$\Delta(t) = \frac{1}{T} \sum_{r=-\infty}^{\infty} e^{ir\omega t}, \qquad \Delta'(t) = \frac{1}{T} \sum_{r=-\infty}^{\infty} (ir\omega) e^{ir\omega t}$$

and so on.

Putting the series in this identity and equating the terms in $e^{ir\omega t}$ (for equal r) we find :

$$\mu_r^* = \mu_r + \frac{1}{T}\left[e^{-ir\omega t_1}\sum_{k=0}^{n-1}\overset{1}{\eta}_k\, m_k(ir\omega) + \right.$$
$$\left. + e^{-ir\omega t_2}\sum_{k=0}^{n-1}\overset{2}{\eta}_k\, m_k(ir\omega) + \ldots\right].$$

Thus equation (5.91) is proved. Then, in order to find μ_r^*, we must find μ_r.

The value of μ_r, the Fourier coefficient of the function $M(P)y$ is computed from the usual formulae for calculating the Fourier coefficients of the expansion of a periodic function in series :

$$\mu_r = \frac{1}{T}\int_{t_0}^{t_2} M(P)\, y e^{-ir\omega t}\, dt =$$
$$= \frac{1}{T}\left[\int_{t_0}^{t_1} M(P)\, y e^{-ir\omega t}\, dt + \int_{t_1}^{t_2} M(P)\, y e^{-ir\omega t}\, dt\right]. \tag{5.92}$$

To calculate the integrals in this expression, we put the value of $\Psi(t)$ in formula (5.88) and, considering only the sections which do not include points of discontinuity, and replacing P^* by P, we obtain :

$$L(P)\, x = M(P)\, y + \Phi(t) + \lambda K(0) - \varkappa D(0)\,. \tag{5.93}$$

In the periodic solution we are considering (see Fig. 214), for $t_0 < t < t_1$ $y = 0$ and for $t_1 < t < t_2$, $x = 0$.

Reference to Fig. 214 indicates that when $y = 0$

$$M(P)\, y = 0, \tag{5.94}$$

$$y(t_1 - 0) = y'(t_1 - 0) = \ldots = y^{(n-1)}(t_1 - 0) = 0\,; \tag{5.95}$$

and that when $x = 0$ we have

$$L(P)\, x = 0, \tag{5.96}$$

$$x(t_2 - 0) = x'(t_2 - 0) = \ldots = x^{(n-1)}(t_2 - 0) = 0. \tag{5.97}$$

The equations (5.95) and (5.97) will be used later. For the present we note that from (5.94) it follows that

$$\int_{t_0}^{t_1} M(P)\, y e^{-ir\omega t}\, \mathrm{d}t = 0.$$

To apply (5.96) we make use of the equation (5.95), which, on using (5.96) reduces to the equation

$$M(P)\, y = -\,\Phi(t) - \lambda K(0) + \varkappa D(0) \quad (t_1 < t < t_2).$$

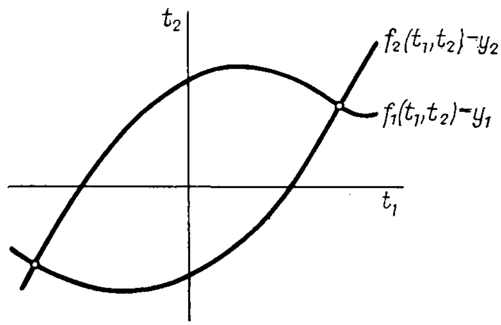


FIG. 216

Therefore the formula (5.92) takes the following form :

$$\mu_r = -\frac{1}{T}\int_{t_1}^{t_2} [\Phi(t) + \lambda K(0) - \varkappa D(0)]\, e^{-ir\omega t}\, \mathrm{d}t,$$

and the coefficients μ_r can be calculated easily from the given function $\Phi(t)$.

Taking (5.90) and (5.91) into account, we obtain :

$$\alpha_r = \frac{\mu_r + \varepsilon_r}{L(ir\omega)} + \frac{1}{TL(ir\omega)} \times$$
$$\times \left[e^{-ir\omega t_1} \sum_{k=0}^{n-1} \overset{1}{\eta}_k m_k(ir\omega) + e^{-ir\omega t_2} \sum_{k=0}^{n-1} \overset{2}{\eta}_k m_k(ir\omega) \right], \qquad (5.98)$$

$$\beta_r = \frac{\mu_r}{M(ir\omega)} + \frac{1}{TM(ir\omega)} \times$$
$$\times \left[e^{-ir\omega t_1} \sum_{k=0}^{n-1} \overset{1}{\eta}_k m_k(ir\omega) + e^{-ir\omega t_2} \sum_{k=1}^{n-1} \overset{2}{\eta}_k m_k(ir\omega) \right].$$

Putting these expressions for a_r and β_r in the Fourier series for x and y and collecting terms with the same η, we find

$$x = \sum_{k=0}^{n-1} [R_k (t - t_1) \overset{1}{\eta}_k + R_k (t - t_2) \overset{2}{\eta}_k] + R(t), \tag{5.99}$$

$$y = \sum_{k=0}^{n-1} [S_k (t - t_1) \overset{1}{\eta}_k + S_k (t - t_2) \overset{2}{\eta}_k] + S(t). \tag{5.100}$$

Here R_k, S_k, R and S are convergent Fourier series :

$$\left.\begin{aligned} R_k &= \frac{1}{T} \sum_{r=-\infty}^{\infty} \frac{m_k(ir\omega)}{L(ir\omega)} e^{ir\omega t}, & S_k &= \frac{1}{T} \sum_{r=-\infty}^{\infty} \frac{m_k(ir\omega)}{M(ir\omega)} e^{ir\omega t}, \\ R &= \sum_{r=-\infty}^{\infty} \frac{\mu_r + \varepsilon_r}{L(ir\omega)} e^{ir\omega t}, & S &= \sum_{r=-\infty}^{\infty} \frac{\mu_r}{M(ir\omega)} e^{ir\omega t}. \end{aligned}\right\} \tag{5.101}$$

Now we return to the conditions (5.95) and (5.97) and require that the functions x and y determined by equations (5.99), (5.100) shall satisfy these conditions. This leads to the following equation :

$$\left.\begin{aligned} &\sum_{k=0}^{n-1} [R_k (t_2 - t_1 - 0) \overset{1}{\eta}_k + R_k (-0) \overset{2}{\eta}_k] + \\ &\qquad + R(t_2 - 0) = 0, \\ &\sum_{k=0}^{n-1} [R'_k (t_2 - t_1 - 0) \overset{1}{\eta}_k + R'_k (-0) \overset{2}{\eta}_k] + \\ &\qquad + R'(t_2 - 0) = 0, \\ &\cdots\cdots\cdots\cdots\cdots \\ &\sum_{k=0}^{n-1} [R_k^{(n-1)} (t_2 - t_1 - 0) \overset{1}{\eta}_h + \\ &+ R_k^{(n-1)}(-0)\overset{2}{\eta}_k] + R^{(n-1)}(t_2 - 0) = 0, \\ &\sum_{k=0}^{n-1} [S_k (-0) \overset{1}{\eta}_k + S_k (t_1 - t_2 - 0) \overset{2}{\eta}_k] + \\ &\qquad + S(t_1 - 0) = 0, \\ &\sum_{k=0}^{n-1} [S'_k (-0) \overset{1}{\eta}_k + S'_k (t_1 - t_2 - 0) \overset{2}{\eta}_k] + \\ &\qquad + S'(t_1 - 0) = 0, \\ &\cdots\cdots\cdots\cdots\cdots \\ &\sum_{k=0}^{n-1} [S_k^{(n-1)} (-0) \overset{1}{\eta}_k + \\ &\qquad + S_k^{(n-1)} (t_1 - t_2 - 0) \overset{2}{\eta}_k] + \\ &\qquad + S^{(n-1)} (t_1 - 0) = 0. \end{aligned}\right\} \tag{5.102}$$

We recall that $\omega = \dfrac{2\pi}{t_2}$ when $\Phi(t) = \text{const.}$ since $t_0 = 0$, and in the case of forced oscillations ω is given. Therefore in either case the equations (5.102) contain only the unknown times t_1 and t_2 apart from the unknowns η which enter linearly in them.

If the quantities t_1 and t_2 are arbitrarily given, and we put them in the system of equations we have just obtained, then we find* from it all the values of the discontinuities $\overset{1}{\eta}_k$ and $\overset{2}{\eta}_k$. Putting these $\overset{1}{\eta}_k$ and $\overset{2}{\eta}_k$ in the series (5.99) and (5.100) we find $x(t)$ and $y(t)$. These $x(t)$ and $y(t)$ will satisfy the equation

$$L(P^*)\,x = M(P^*)\,y + \Psi(t).$$

During the time $t_0 < t < t_1$, the motion will be along the x-axis (i.e. $y = 0$) and during the time $t_1 < t < t_2$ it will be along the y-axis (i.e. $x = 0$), but it cannot be ensured that the crossing from the x-axis to the y-axis is from the point Q_1 to the point Q_1', or that the crossing from the y-axis to the x-axis is from the point Q_2 to the point Q_2'. To ensure this "descent condition"** it is sufficient that the following two conditions are satisfied (see Fig. 213) :

$$y(t_1 + 0) = y_1, \quad y(t_2 - 0) = y_2,$$

where

$$y_1 = \frac{\sigma_1 - \dfrac{h_1 - h_2}{k_2 - k_1}}{\beta}, \qquad y_2 = \frac{\sigma_2 - \dfrac{h_1 - h_2}{k_2 - k_1}}{\beta} \tag{5.103}$$

These conditions fix the position of the points Q_1' and Q_2, and therefore also the points Q_1 and Q_2', since the directions of the jumps from Q_1 to Q_1' and from Q_2 to Q_2' are given by the chosen linear transformation (5.87).

* It is assumed that for these t_1 and t_2 the determinant of the system is not zero. For the case when this determinant is equal to zero (see p. 426).

** In the theory of relay systems (Section 9) the condition 2° (see p. 393) corresponds to these conditions.

Because of the fact that $\overset{1}{\eta}_0 = y(t_1 + 0) - y(t_1 - 0)$ and $\overset{2}{\eta}_0 = y(t_2 + 0) - y(t_2 - 0)$ while $y(t_1 - 0) = y(t_2 + 0) = 0$ (the motion is along the x-axis) the condition (5.103) can be rewritten as :

$$\overset{1}{\eta}_0 = y_1, \quad \overset{2}{\eta}_0 = -y_2. \tag{5.104}$$

Assuming for the time being that the determinant of the system (5.102) $\Delta(t_1, t_2) \neq 0$ * we solve it with respect to $\overset{1}{\eta}_0$ and $\overset{2}{\eta}_0$. Let this give $\overset{1}{\eta}_0 = f_1(t_1, t_2)$, $\overset{2}{\eta}_0 = f_2(t_1, t_2)$. Then the conditions

$$f_1(t_1, t_2) = y_1, \quad f_2(t_1, t_2) = -y_2 \tag{5.105}$$

can be used to determine t_1 and t_2.

For example, in the t_1, t_2-plane we can construct the curves (5.105) by the usual method, and find their points of intersection (Fig. 216). Thus values of t_1 and t_2 corresponding to these points are "pretenders to the periodic solutions". They in fact determine the periodic solutions if the following conditions are satisfied :**

(a) The inequality $0 < t_1 < t_2$ (for auto-oscillations) or $t_2 - T < t_1 < t_2$ (for forced oscillations) is satisfied.

(b) There is no "switching" during a period when t is different from t_1 and t_2.

To verify that condition (a) is satisfied it is convenient to construct in the (t_1, t_2)-plane the angle between the t_2-axis and the bisector in the first quadrant (Fig. 217) in the auto-oscillatory case, and the strip between the bisector and the straight line $t_2 - t_1 = T$ in the case of forced oscillations (Fig. 218). The points of intersection of these curves lying in the unshaded region (Fig. 217 in the auto-oscillatory case and Fig. 218 in the case of forced oscillations) must be rejected. In order to verify that condition (b) is satisfied by the remaining points of intersection it is necessary, with the obtained t_1 and t_2, to find all the η from the system (5.102), to put them in (5.99) and

* See the previous footnote but one.

** In the theory of relay systems (Section 9) condition 3° (see p. 393) corresponds to these conditions.

(5.100) and to construct the obtained periodic solution at the ends of the period.

Let $x_1 = \dfrac{\sigma_1 - \dfrac{h_1 - h_2}{k_2 - k_1}}{\alpha}$ be the abscissa of the point Q_1 and

$y_2 = \dfrac{\sigma_2 - \dfrac{h_1 - h_2}{k_2 - k_1}}{\beta}$ the ordinate of the point Q_2 (see Fig. 213).

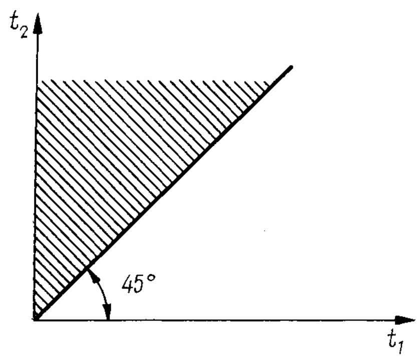


FIG. 217

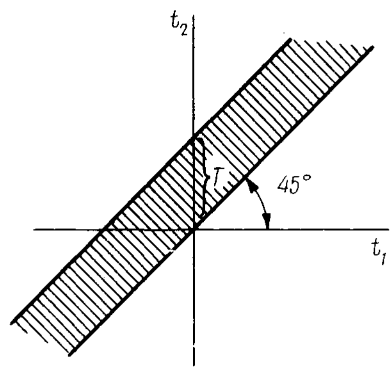


FIG. 218

If in the constructed solution the value of x_1 is attained for any t different from t_1, or the value of y_2 is attained for any t different from t_2, then this periodic solution does not satisfy condition (b) and must be rejected.

All the pairs of numbers t_1, t_2 satisfying the period equation (5.105) and the conditions (a) and (b), determine periodic solutions which can be constructed according to the given formulae. In doing this we can sometimes omit some periodic solutions, since we have temporarily assumed that $\Delta (t_1, t_2) \neq 0$.

When $\Delta (t_1, t_2) = 0$ the system (5.102), generally speaking, is a contradiction, and there are no periodic solutions. But in exceptional cases, for a particular choice of the free terms, the system (5.102) may be indeterminate and its solutions may contain some for which $\overset{1}{\eta}_0 = y_1$, $\overset{2}{\eta}_0 = -y_2$. These cases are isolated by the usual investigation of the system of algebraic equations (5.102).

The calculation of the coefficients of the equations (5.102)

From what has been said it follows that the determination of the periodic solutions of equation (5.86) reduces to the repeated solution of the system of linear algebraic equations (5.102). The values of the functions $R(t)$, $S(t)$, $R_k(t)$ and $S_k(t)$ and their derivatives up to and including the $(n-1)$th order at the fixed moment of time are the coefficients of these equations. The values of the functions R, S, R_k, S_k themselves can be determined from their Fourier expansion (5.101), since these series converge. But difficulties arise in determining the values of their derivatives. If we attempt to find them by the term-by-term differentiation of the series (5.101) then we must bear in mind that the convergence of the series obtained after the rth differentiation becomes rapidly worse as r increases, and from some $r \leqslant n-1$ these series can turn out to be divergent.

In order to avoid these difficulties, we can make use of the method for improving the convergence of series which was devised by A. N. Krylov. We explain the use of this method with the example of the function

$$R_k = \frac{1}{T} \sum_{r=-\infty}^{\infty} \frac{m_k(ir\omega)}{L(ir\omega)} e^{ir\omega t},$$

where

$$m_k(S) = b_{n-k-1} + b_{n-k-2} S + \ldots + b_0 S^{n-k-1}.$$

Let l be an arbitrary whole positive number. We now divide the polynomial $S^{k+l} m_k(S)$ by $L(S)$. We denote the quotient by $\gamma_0 S^{l-1} + \gamma_1 S^{l-2} + \ldots + \gamma_{l-1}$ and the remainder by $c_0 S^{n-1} + c_1 S^{n-2} + \ldots + c_{n-1}$. Then

$$S^{k+l} m_k(S) = (\gamma_0 S^{l-1} + \gamma_1 S^{l-2} + \ldots + \gamma_{l-1}) L(S) + c_0 S^{n-1} + c_1 S^{n-2} + \ldots + c_{n-1},$$

which gives

$$\frac{m_k(S)}{L(S)} = \frac{\gamma_0}{S^{k+1}} + \frac{\gamma_1}{S^{k+2}} + \ldots + \frac{\gamma_{l-1}}{S^{k+l}} + \frac{c_0 S^{n-1} + c_1 S^{n-2} + \ldots c_{n-1}}{S^{k+l} L(S)}. \tag{5.106}$$

Using the identify (5.106), we replace the Fourier series for R_k by a sum of Fourier series*

$$R_k = \frac{b_{n-k-1}}{a_n} + \gamma_0 H_{k+1} + \gamma_1 H_{k+2}(t) + \dots \\ \dots + \gamma_{l-1} H_{k+l}(t) + R_k^*(t), \tag{5.107}$$

where

$$H_j(t) = \frac{1}{T} \sum_{r=-\infty}^{\infty}{}' \frac{1}{(ir\omega)^j} e^{ir\omega t} \tag{5.108}$$

$$(j = k+1, k+2, \dots, k+l),$$

$$R_k^*(t) = \frac{1}{T} \sum_{r=-\infty}^{\infty}{}' \frac{c_0 (ir\omega)^{n-1} + c_2 (ir\omega)^{n-2} + \dots + c_n}{(ir\omega)^{k+l} L(ir\omega)} e^{ir\omega t} \tag{5.109}$$

and Σ' denotes the sum in which the term for $r = 0$ is omitted.

In contrast to the series for $R_k(t)$, the series for $R_k^*(t)$, when $l \geqslant n - k - 1$ possesses the property that all the series obtained from it by term-by-term differentiation, up to and including the $(n - 1)$th order, converge. Here the larger l is, the better will the convergence of the obtained series be.

We can therefore select the number l to be so large that the values of $R_k^*(t)$ and all its derivatives up to the (n-1) th inclusively can be calculated from the first few harmionics.

We now consider the function. $H\ (t)$. We note that

$$H_j(t) = \frac{1}{\omega^j T} h_j(\omega t), \text{ where } h_j(z) = \sum_{r=-\infty}^{\infty}{}' \frac{1}{(ir)^j} e^{irz}.$$

Differentiating $h_j(z)$ term-by-term, we obtain :

$$h_j'(z) = \sum_{r=-\infty}^{\infty}{}' \frac{1}{(ir)^{j-1}} e^{irz} = h_{j-1}(z), \tag{5.110}$$

i.e.

$$h_j(z) = \int h_{j-1}(z)\, dz + C. \tag{5.111}$$

* The identity (5.106) for $S = ir\omega$ applies to any r except $r = 0$. Theretore in the expression for R_k the constart term of the series (corresponding јo $r = 0$) is taken separately.

In this recurrence formula the constant of integration is determined from the conditions

(a) when j is even $\int_0^\pi h_j(z)\,dz = 0$;

(b) when $j > 1$ and odd, $h_j(0) = 0$.

The first condition follows from the fact that h_j (for even j) is an even function, whose Fourier expansion has no free term; the second condition is a consequence of the fact that h_j is a continuous odd function when j is odd and larger than one.

We note that

$$h_1(z) = \sum_{r=-\infty}^{\infty}{}' \frac{1}{(ir)} e^{irz} = 2 \sum_{n=1}^{\infty} \frac{\sin nz}{n}$$

is a periodic function of period 2π which, for $0 < z < 2\pi$, is given by the formula

$$h_1(z) = \pi - z \quad (0 < z < 2\pi). \tag{5.112}$$

Putting this value of $h_1(z)$ in (5.111) we obtain:

$$h_2(z) = \int (\pi - z)\,dz + C = \pi z - \frac{z^2}{2} + C.$$

We find the quantity C from the condition

$$\int_0^\pi h_2(z)\,dz = \left(\pi \frac{z^2}{2} - \frac{z^3}{6} + Cz\right)\Big|_0^\pi = 0,$$

i.e.

$$C = -\frac{\pi^2}{3} \quad \text{and} \quad h_2(z) = \pi z - \frac{z^2}{2} - \frac{\pi^2}{3}.$$

Then, putting this $h_2(z)$ in (5.111), we find $h_3(z)$, and so on. We derive standard formulae for h from h_1 to h_{11}:

$$h_1 = -z + \pi,$$

$$h_2 = -\frac{z^2}{2} + \pi z - \frac{\pi^2}{3},$$

$$h_3 = -\frac{z^3}{6} + \frac{\pi z^2}{2} - \frac{\pi^2 z}{3},$$

$$h_4 = -\frac{z^4}{24} + \frac{\pi z^3}{6} - \frac{\pi^2 z^2}{6} + \frac{\pi^3}{45},$$

$$h_5 = -\frac{z^5}{120} + \frac{\pi z^4}{24} - \frac{\pi^2 z^3}{18} + \frac{\pi^4 z}{45},$$

$$h_6 = -\frac{z^6}{720} + \frac{\pi z^5}{120} - \frac{\pi^2 z^4}{72} + \frac{\pi^4 z^2}{90} - \frac{2\pi^6}{945},$$

$$h_7 = -\frac{z^7}{5040} + \frac{\pi z^6}{720} - \frac{\pi^2 z^5}{360} + \frac{\pi^4 z^3}{270} - \frac{2\pi^6 z}{945},$$

$$h_8 = -\frac{z^8}{40\,320} + \frac{\pi z^7}{5040} - \frac{\pi^2 z^6}{2160} + \frac{\pi^4 z^4}{1080} - \frac{\pi^6 z^2}{945} + \frac{23\pi^8}{37\,800},$$

$$h_9 = -\frac{z^9}{362\,880} + \frac{\pi z^8}{40\,320} - \frac{\pi^2 z^7}{15\,120} + \frac{\pi^4 z^5}{5400} - \frac{\pi^6 z^3}{2835} + \frac{23\pi^8 z}{37\,800},$$

$$h_{10} = -\frac{z^{10}}{3\,628\,800} + \frac{\pi z^9}{362\,880} - \frac{\pi^2 z^8}{120\,960} + \frac{\pi^4 z^7}{32\,400} - \frac{\pi^6 z^4}{11\,340} + \frac{23\pi^8 z^2}{75\,600} - \frac{5813\,\pi^{10}}{19\,958\,400},$$

$$h_{11} = -\frac{z^{11}}{39\,916\,800} + \frac{\pi z^{10}}{3\,622\,800} - \frac{\pi^2 z^9}{1\,088\,640} + \frac{\pi^4 z^6}{226\,800} - \frac{\pi^6 z^5}{56\,700} + \frac{23\pi^9 z^3}{226\,800} - \frac{5813\pi^{10} z}{19\,958\,400}.$$

On the strength of (5.110) we have

$$\frac{d^s h_j}{dz^s} = h_{j-s}(z). \tag{5.113}$$

Using (5.113) and (5.107) we obtain :

$$R_k(t) = \frac{b_{n-k-1}}{a_n} + \frac{1}{2\pi}\left[\frac{\gamma_0}{\omega^k} h_{k+1}(\omega t) + \frac{\gamma_1}{\omega^{k+1}} h_{k+2}(\omega t) + \dots + \frac{\gamma_{l-1}}{\omega^{l+k-1}} h_{k+1}(\omega t)\right] + R_k^*(t). \tag{5.114}$$

$$\frac{d^s R_k(t)}{dt^s} = \frac{1}{2\pi}\left[\frac{\gamma_0}{\omega^k} h_{k+1-s}(\omega t) + \frac{\gamma_1}{\omega^{k+1}} h_{k+2-s}(\omega t) + \dots + \frac{\gamma_{l-1}}{\omega^{l+k-1}} h_{k+l-s}(\omega t)\right] + \frac{d^s R_k^*(t)}{dt^s},$$

where $h_g = 0$ if $g < 0$ and $R_k^*(t)$ and its derivatives are determined from (5.109).

Thus, the calculation of the values of R_k or $R_k^{(s)}$ reduces to the summation of the values of the first few harmonics of the rapidly converging series for $R_k^*(t)\left(\text{or } \dfrac{\mathrm{d}^s R_k^*(t)}{\mathrm{d}t^s}\right)$. In an exactly similar way we can form formulae for the values of S and S_k and their derivatives*.

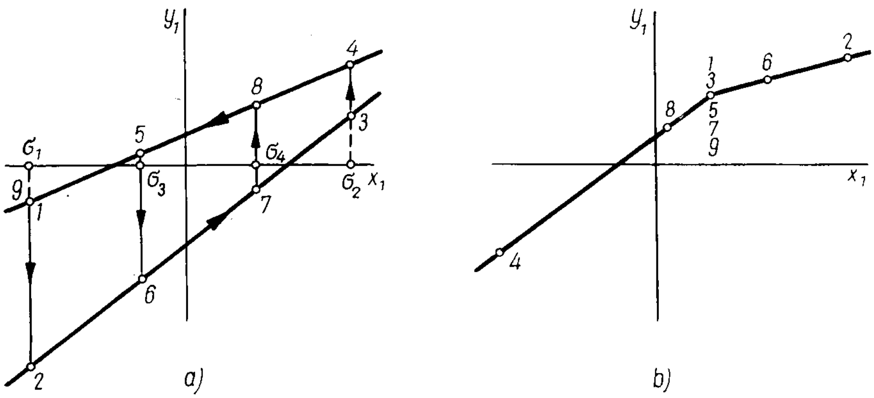


FIG. 219

Generalized problems

We can now generalize the class of problems, solved by this method, in three directions.

1. We can consider characteristics formed of more than two segments of the same two straight lines, i.e. we can determine not only the simplest, but also more complicated states in systems with a two-stage broken line characteristic. Examples of this sort are shown in Fig. 219, where during each period the numbered points occur in numerical order.

2. We can consider continuous or discontinuous characteristics, formed from any number of segments of various straight lines, i.e. provided the slopes of the lines are equal to one of the two given numbers k' and k''. Examples of such characteristics are given in Fig. 220.

* We note for this that for the function S_k the coefficient $\gamma_0 = 1$ and $\gamma_1 = \ldots = \gamma_{n-k-1} = 0$.

3. We can consider more general ways of crossing from one stage of the characteristic to the other. Thus, when there is clearance or dry friction in an inertialess element for example (Fig. 221), the coordinate y_1 either changes with x_1 ($y_1 = x_1 +$ const) if the clearance is "selected", or does not change ($y_1 =$ const), if the clearance is "self-selected". The characteristic consists of segments of straight

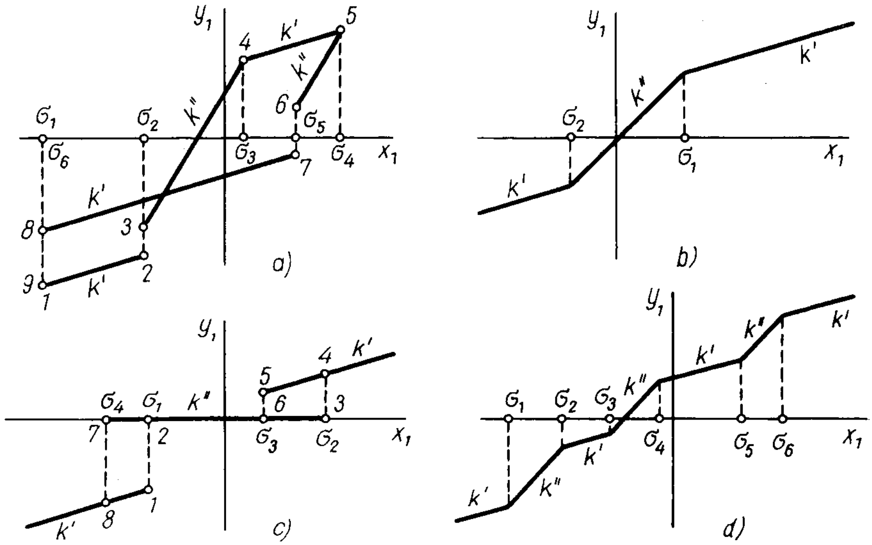


FIG. 220

lines of slope 1 or 0 ; the crossing from the first line to the second occurs at the time when $\dot{x}_1$ first becomes zero, whatever $x_1 = \bar{x}_1$ is at this time, and the second crossing on the sloping line occurs at the time when x_1 first differs from $\bar{x}_1$ by a given amount.

We can show other examples of problems for which the characteristic consists of straight lines parallel to two given lines where the the conditions of crossing from one segment to another are determined not by the equations $x_1 = \sigma_1$, but by other (not necessarily linear) relations between x_1 and its derivatives at various times.

Bearing in mind this class of problems with all these types, we return to the derived equation

$$D(P^*)\, x_1 = K(P^*)\, y_1 + \Phi\,(t), \qquad (5.115)$$

$$y_1 = f(x_1),$$

obtained by eliminating all the x_j except x_1 from the system (5.85). We shall now assume that the period T can be divided into N stages, occurring at times $t_1 - t_0$, $t_2 - t_1$, . . . , $t_N - t_{N-1}$ (here $t_N - t_0 = T$), and that during each step the point x_1, y_1 moves along a straight line having one of two given directions, characterised by the slopes k' and k''.

We shall assume that for each ith step ($i = 1, 2, \ldots, N$) we are given:

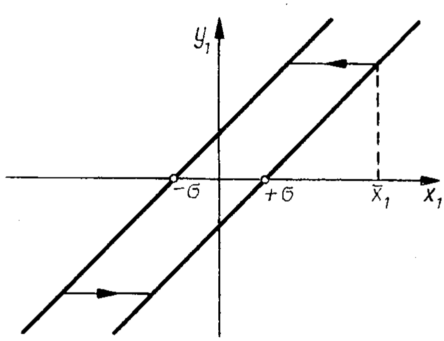


FIG. 221

(a) the equation of the straight line along which the point x_1, y_1 moves during this step, in the form

$$y_1 = k_i x_1 + \varphi_1 x_{l,i-l} + \psi_i y_{l,i-l} + \theta_i, \tag{5.116}$$

where k_i is one of the numbers k_i and k'', and φ_i, ψ_i and θ_i are given numbers, and x_{1i-1}, $y_{1,\, i-1}$ is the point at which the previous $(i-1)$th, step ended, i.e.

$$x_{l,i-l} = x_1(t_{i-l} - 0), \quad y_{l,i-} = y_1(t_{i-l} - 0);$$

(b) the end conditions of the step (the "descent conditions") in the form of a linear relation between $x_1(t)$ and its derivatives at the times t_i and t_{i-1}:

$$\sum_{j=1}^{n} \zeta_j x_1^{(j-1)}(t_i) + \sum_{j=1}^{n} \pi_j x_1^{(j-1)}(t_{i-1}) + \zeta = 0. \tag{5.117}$$

At time t_i, when relation (5.117) is first satisfied, the ith step is concluded.

Due to the fact that the degree of $K(P^*)$ in (5.115) is less than that of $D(P^*)$, the function x_1 is continuous and there is no need to write out conditions for the beginning of each step — they are determined from the "descent" conditions by the continuity of x_1.

We illustrate what we have said by the following examples.

Example 1. The characteristic $f(x_1)$ is a three-stage broken line with parallel outer stages (see Fig. 220b). We look for the continuous simplest periodic state.

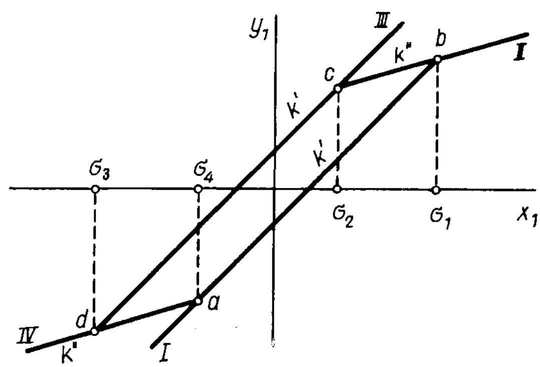


Fig. 222

In this case $N = 4$, for all steps $\varphi_i = \psi_i = 0$, for the middle stage $k = k''$, $\theta = 0$; for the outer stages $k = k'$ and θ is equal to $(k' - k'')\,\sigma_1$ for the right-hand stage and to $(k' - k'')\,\sigma_2$ for the left-hand stage. The equations $x_1 = \sigma_1$ and $x_1 = \sigma_2$ are respectively the end conditions of the step.

Example 2. The same characteristic with a loop (Fig. 222). If at time $t = t_0$ the points is at a, and then moves along the line I (in particular, away from a) until such time as the point b is reached first, then along the line II until the point c is reached for the first time, and so on then in this case also $N = 4$, and the data for each step is written out in Table XVII.

Example 3. This system differs from a linear system by the presence of an inertialess element in which clearance exists or dry (coulomb) friction acts.

For this example the values of k, φ, ψ, θ and ξ are given in Table XVIII for all steps.

We now perform the same substitution in equation (5.115) as we made in (5.87) with the coefficients determined by equation

(5.89). The values of h_1 and h_2 entering in (5.89) are taken as any two arbitrary stages of the characteristic, having respectively the slopes k' and k''. As a result of the transformation the first of these straight lines is transformed in the x, y-plane into the x-axis, and the second into the y-axis. With a linear transformation the parallelism of the straight lines is preserved. Therefore all the straight lines with slope k' transform into lines parallel to the x-axis in the $x\,y$-plane, and all the lines with slope k'' into lines parallel* to the y-axis.

TABLE XVII

Stage of motion of point	Equation of line	Descent condition
From a to b	$y_1 = k'x_1 + (h_1 - k'\sigma_1)$, i.e. $k_1 = k'$, $\varphi_1 = \psi_1 = 0$, $\theta_1 = h_1 - k'\sigma_1$,	$x_1 = \sigma_1$
From b to c	$y_1 = k''x_1 + (h_1 - k''\sigma_1)$, i.e. $k_2 = k''$, $\varphi_2 = \psi_2 = 0$, $\theta_2 = h_1 - k'\sigma_1$,	$x_1 = \sigma_2$
From c to d	$y_1 = k'x_1 + [h_1 + (k'' - k')\sigma_2 - k''\sigma_1]$, i.e. $k_3 = k'$, $\varphi_3 = \psi_3 = 0$, $\theta_3 = h_1 + (k'' - k')\sigma_2 - k''\sigma_1$,	$x_1 = \sigma_3$
From d to a	$y_1 = k''x_1 + [(k' - k'')\sigma_4 + h_1 - k''\sigma_1 + (k'' - k'\sigma_2)]$, i.e. $k_4 = k''$, $\varphi_4 = \psi_4 = 0$, $\theta_4 = (k' - k'')\sigma_3 + h_1 - k''\sigma_1 + (k'' - k')\sigma_2$	$x_1 = \sigma_4$

The equation (5.115) becomes

$$L(P^*)\,x = M(P^*)\,y + \Psi(t), \tag{5.118}$$

where $\Psi(t)$ is a Fourier series which differs from the series $\Phi(t)$ only in the constant term.

* The need for ensuring this also forces us to restrict ourselves to characteristics consisting of stages parallel to two given straight lines. If the characteristic contained stages with more than two different slopes, then the transformed characteristic would also contain links not parallel to the x- and y-axes, and all the subsequent reasoning (in particular, the calculation of μ) could no longer be made by this method.

TABLE XVIII

No. of stage	k	φ	ψ	θ	Descent condition
I	1	0	0	$-\sigma$	$\dot{x}(t_1)=0$
II	0	0	1	0	$x(t_2)-x(t_1)=-2\sigma$
III	1	0	0	σ	$\dot{x}(t_3)=0$
IV	0	0	1	0	$x(t_4)-x(t_3)=2\sigma$

The descent conditions (5.117), as a result of this transformation, become

$$\sum_{j=1}^{n}\zeta_j\,[ax^{(j-1)}(t_i)+\beta y^{(j-1)}(t_i)]+$$
$$+\sum_{j=1}^{n}\pi_j\,[ax^{(j-1)}(t_{i-1})+\beta y^{(j-1)}(t_{i-1})]+\zeta+\varkappa\,(\zeta_1+\pi)=0 \tag{5.119}$$
$$(i=1,2,\ldots,N).$$

The motion of the point in the x, y-plane during the periodic state can be pictured in the following way (example in Fig. 223). At time t_0 the point is on the straight line I (on the x-axis) and moves along it until at time t_1 it reaches point 2, where the first of the conditions (5.119) is first satisfied. If the line II passes through this point 2, then the motion will continue along it; if it does not pass through it, then the point jumps instantaneously to the straight line II via a line with slope $\tan\varphi=\dfrac{\beta}{a}$. Then the point moves along the line II until time t_2 when it reaches the point 4, where the second of the conditions (5.119) is first satisfied, and so on. Due to the fact that the motion in the x_1, y-plane occurs only along a straight line parallel to one of the axes, in each interval either x or y is constant. Figure 224 gives an example of the change in x and y corresponding to the characteristic and sequence of switching shown in Fig. 223.

If we keep the old notation of ε_r for the Fourier coefficients of the series $\psi\,(t)$ and if, as before, we look for periodic solutions in the form of Fourier series with Fourier coefficients a_r and β_r, , then the relations (5.90) connecting a_r and β_r with the Fourier coefficients μ_r^* of the function $M(P^*)\,y$ remain absolutely in force.

We retain the notation μ_r for the Fourier coefficients of the periodic function $M(P)\,y$. The formula (5.91) connecting μ_r^* and μ_r can be used in this more general case also, but we must retain not two, but N terms of this formula — each term relates respectively to one of the times $t_1, t_2, \ldots, t_N$. Therefore now formula (5.91) contains Nn unknowns $\mathfrak{z}_j^i$, the discontinuities of the function y and its derivatives up to the $(n-1)$th at the times $t_1, t_2, \ldots, t_N$. All these unknown discontinuities enter linearly in (5.91), but the unknowns $t_1, t_2, \ldots, t_N$ enter non-linearly.

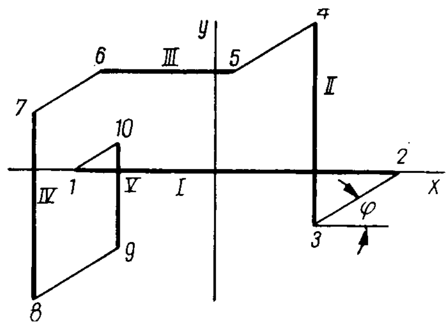


FIG. 223

The calculation of the values of μ_r is carried out in general terms just as in the simplest case considered above, since in the expression

$$\mu_r = \frac{1}{T}\int_{t_0}^{t_N} M(P)\,y e^{-ir\omega t}\,\mathrm{d}t \tag{5.120}$$

the integral can be expressed in terms of $t_1, t_2, \ldots, t_N$ if we divide it into the N integrals

$$\mu_r = \frac{1}{T}\left[\int_{t_0}^{t_1} M(P)\,y e^{-ir\omega}\,\mathrm{d}t + \int_{t_1}^{t_2} M(P)\,y e^{-ir\omega t}\,\mathrm{d}t + \ldots + \int_{t_{N-1}}^{t_N} M(P)\,y e^{-ir\omega t}\,\mathrm{d}t\right] \tag{5.121}$$

and if we make use of the fact that between the limits of integration of each integral either x or y is constant.

Indeed, if in the interval $t_{i-1} < t < t_i$ we keep $y = \text{const}_i$, then this condition can be replaced by the conditions

$$M(P)\,y = b_n \,\text{const}_i, \tag{5.122}$$

$$\begin{gathered} y\,(t_i - 0) = \text{const}_i, \quad y'\,(t_i - 0) = y''\,(t_i - 0) = \ldots \\ \ldots = y^{(n-1)}\,(t_i - 0) = 0. \end{gathered} \tag{5.123}$$

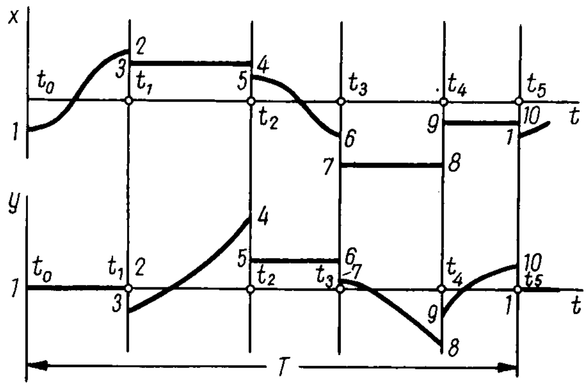


Fig. 224

Using (5.122) we calculate the corresponding integral in (5.121) :

$$\int\limits_{t_{i-1}}^{t_i} M(P)\,ye^{-ir\omega t}\,dt = b_n \,\text{const}_i \int\limits_{t_{i-1}}^{t_i} e^{-ir\omega t}\,dt =$$

$$= -\frac{b_n \,\text{const}_i}{ir\omega}\left[e^{-ir\omega t_i} - e^{-ir\omega t_{i-1}}\right].$$

If in the interval $t_{j-1} < t < t_j$ $x = \text{const}_j$; then for this interval

$$L(P)\,x = a_n \,\text{const}_j, \tag{5.124}$$

$$\begin{gathered} x\,(t_j - 0) = \text{const}_j, \quad x'\,(t_j - 0) = x''\,(t_j - 0) = \ldots \\ \ldots = x^{(n-1)}\,(t_j - 0) = 0. \end{gathered} \tag{5.125}$$

In the intervals not containing points of discontinuity, in formula (5.118) we can replace P^* by P. Then from (5.124) and (5.118) it follows that

$$M(P)\,y = a_n \,\text{const}_j - \Psi(t)$$

and the corresponding integral in (5.121) is

$$\int_{t_{j-1}}^{t_j} M(P)\, y e^{-ir\omega t}\, \mathrm{d}t = -\frac{a_n \operatorname{const}_j}{ir\omega}\left[e^{-ir\omega t_j} - e^{-ir\omega t_{j-1}}\right] - \\ - \int_{t_{j-1}}^{t_j} \Psi(t)\, e^{-ir\omega t}\, \mathrm{d}t$$

which is easily calculated, since the function $\Psi(t)$ is given.

In this way all the integrals in (5.121) are calculated, giving the value of μ_r, but the Nn initial conditions of the form (5.123) or (5.125) remain unused.

Putting the resulting value of μ_r in (5.91) we find μ_r^*, and putting this in (5.90) we express all the α_r and β_r in terms of the linearly entering unknowns $\mathfrak{z}_i^j$ and in terms of the N non-linearly entering unknowns $t_1, t_2, \ldots, t_N$.

Now requiring that the Fourier series for $x(t)$ and $y(t)$ obtained in this way shall satisfy all Nn conditions of the form (5.123) or (5.125), we obtain a non-homogeneous system of linear algebraic equations with respect to the unknowns μ_i^j. The coefficients in these linear algebraic equations depend on the required unknowns $t_1, t_2, \ldots, t_N$. Taking, for the time being, the determinant of this system as different from zero, and solving it, we find all the $\mathfrak{z}_i^j$ as functions of $t_1, t_2, \ldots, t_N$:

$$\mathfrak{z}_i^j = f_i^j(t_1, t_2, \ldots, t_N)\left(\begin{array}{l} i = 0,\ 1, \ldots,\ n-1, \\ j = 1,\ 2, \ldots,\ N \end{array}\right). \tag{5.126}$$

Putting these η_i^j in the expressions for α_r and β_r we find immediately:

$$x(t, t_1, t_2, \ldots, t_N) \text{ and } y(t, t_1, \ldots, t_N). \tag{5.127}$$

We have still not used the "descent conditions" (5.119). Putting (5.127) in (5.119), we obtain a system of N transcendental equations in the required N unknowns $t_1, t_2, \ldots, t_N$ i. e. the period equations.

The required periodic states are found from the solutions of the period equations which satisfy the following conditions:

(a) The inequality

$$t_0 < t_1 < t_2 < \ldots < t_N.$$

with $t_0 = 0$ for auto-oscillations, and $t_0 = T - t_N$ for forced oscillations, must be satisfied.

(b) There is no "switching" during a period, i.e. at a time t which is different from $t_1, t_2, \ldots, t_N$.

All that we said above about the necessity for checking that a zero value of the determinant of the system of algebraic equations does not introduce additional solutions, and everything involved in improving the convergence of the Fourier series, can be extended completely to this more general problem.

We note in conclusion that the number of time unknowns and the corresponding number of period equations which must be formed is halved if the state is symmetric, i.e. if it is known beforehand that

$$x(t) = -x\left(t + \frac{T}{2}\right), \qquad y(t) = -y\left(t + \frac{T}{2}\right).$$

In this case instead of the period T we can restrict ourselves to considering the half-period $\frac{T}{2}$.

In Section 9 this fact has already been used as applied to a relay system. There we determined a symmetric periodic state having two relay switchings per period, but with only one unknown — the period T. The second switching of the relay, due to the symmetry of the state, took place at $\frac{T}{2}$. Moreover, in the case of a relay, the coefficients were determined as functions of T directly, and this considerably simplified the calculation.

We can now summarize the methods we have described in Sections 3—10 for determining the periodic states. Even in the case of piecewise-linear characteristics exact methods are difficult to apply. This difficulty arises from the necessity of solving a system of transcendental equations ; the actual formation of these equations is a relatively simple problem. Only for the very special case of the most simple symmetric states for a symmetric relay characteristic is there only one period equation, which is not difficult to solve. But, on the other hand, the results obtained by exact methods are genuine, and they enable us to determine all the periodic solutions.

Approximate methods are considerably more simple, but even then only in the case of astatic systems and symmetric characteristics.

In other cases approximate methods also lead to the need for solving a system of transcendental equations [such as (5.8) and (5.8′)], although these are usually simpler than the exact period equations. But, on the other hand, approximate methods allow us to find the periodic solutions only when there is a filter or auto-resonance, and even then only some of the periodic solutions can be found : when there is auto-resonance we can only find those solutions with a frequency which is near the auto-resonance frequency, and when there is a filter, only those with a frequency less than three times the cut-off frequency. Approximate methods, therefore, cannot give authentic data about the non-existence of periodic solutions.

11. The Stability of Periodic States Found Exactly

In the beginning of Section 7 it was shown that when the characteristic $f(x_1)$ was smooth, the question of the stability of the periodic state reduced to the investigation of the stability of the equilibrium in the system described by linear equations with periodic coefficients. But here we consider a piecewise-linear characteristic $f(x_1)$ having discontinuities and breaks. In this case also, to reduce the question of the stability of the equilibrium in a linear system with periodic coefficients, we make use of a theorem from the theory of the stability of motions, which is quoted below without proof.*

In this theorem the term "asymptotic stability" is used. When speaking of stability above (see Chapter II) we have always had in mind only the following property : after sufficiently small initial deviations the considered state is restored after a time, i. e. the deviations from the investigated state tend to zero as $t \to \infty$. In essence, the discussion was about convergence in the sense given to this term by the engineer rather than stability as this term is understood in mathematics and technics. We shall now say that the motion is asymptotically stable if there is convergence, i.e. small deviations tend to zero, and if during the setting-up time the deviations from the given state do not go outside a small neighbourhood around the given

* The proof of this theorem is contained in: Aizerman, M. A. and Gantmakher, F. R. "The Stability Of a System of Differential Equations with Discontinuous Right-Hand Sides as Indicated by the Linear Approximation of a Periodic Solution", *Prikl. mat. i mekh.*, Vol XVII, No. 5 (1957).

motion.** Thus, the requirement of asymptotic stability is stricter than the requirement of convergence, and if there is asymtotic stability, then there is certainly convergence.

Proposing to return to (5.1) later, let us first consider the more general equation

$$x_i = f_i(x_1, \ldots, x_n, t) \qquad i = 1, 2, \ldots, n, \tag{5.128}$$

where all the f_i are functions which are periodic in t with a total period τ (in particular, they need not depend on t).

Let the space of $x_1, \ldots, x_n, t$ be intersected by the surfaces $F_\alpha(x_1, \ldots, x_n, t) = 0$ in the regions H_α ($\alpha = 1, 2, \ldots$) It is assumed that each region H_α is given its own functions f_i, sufficiently smooth in this region.

On passing across the surface F_α there can be discontinuities both in the functions f_i themselves, and in their partial derivatives. The integral curves of the equations (5.128) are continuous, although there are breaks on the surfaces F_α.

Suppose, further, that $x_i = \tilde{x}_i(t)$ is a periodic solution of the system (5.128) intersecting the surfaces $F_\alpha = 0$ at the time $t = t_\alpha$.

Together with the system (5.128) we consider the linear equations whose right-hand sides are given in all regions H_α:

$$\Delta\dot{x}_i = \sum_{j=1}^{n} b_{ij}\,\Delta x_j, \quad i = 1, 2, \ldots, n, \tag{5.129}$$

where $b_{ij} = \left[\dfrac{\partial f_i}{\partial x_j}\right]_{x=\tilde{x}(t)}$ are periodic coefficients with a total period τ, defined for any t except $t = t_\alpha$.

To equation (5.129) we add linear relations defining the discontinuities Δx for each instant $t = t_\alpha$:

$$\begin{aligned} \Delta x_i(t_\alpha + 0) - \Delta x_i(t_\alpha - 0) &= \xi_i \sum_{k=1}^{n} h_k^- \,\Delta x_k(t_\alpha - 0) = \\ &= \xi_i \sum_{k=1}^{n} h_k^+ \,\Delta x_k(t_\alpha + 0), \end{aligned} \tag{5.130}$$

** The term "stability" (as understood by Lyapunov) is not used in this book, but the term "asymptotic stability" is introduced for the time being, but will not be used later. Therefore we are not here using the usual strict ε — the definition of these terms.

where

$$h_k^{\pm} = \left[\frac{\dfrac{\partial F_a}{\partial x_k}}{\left(\dfrac{\partial F_a}{dt} \right)^{\pm}} \right]_{x=\tilde{x}(t_a)}$$

Equations (5.129) and the relations (5.130) determine jointly the integral curves which are discontinuous (for $t = t_a$). We call the aggregate of (5.129) and (5.130) the *linear approximation of the equations (5.128) for the periodics solutions* $x_i = \tilde{x}_i\,(t)$. Then if several additional limitations* are laid on the functions f_i and F_a, we have the following theorem.

If the zero solution $\Delta\, x_i = 0$ *of a system of linear approximations (5.129) and (5.130) is asymptotically stable, then the periodic solution of the system of equations (5.128) is also asymptotically stable.*

We return now to the system (5.1).

Suppose, to be specific, that the crossing from one straight line of the characteristic $f(x_1)$ to the other is made at the instant when x_1 first acquires the given values σ_a ($a = 1, 2, \ldots, r$). It is assumed that the periodic solution of the equations (5.128) of period T has been found in the way, for example, previously described in Sections 9 and 10. If each stage of the characteristic has the equation

$$f(x_1) = K_a\, x_1 + S_a\,,$$

then for the system (5.1) equation (5.129) can be rewritten :

$$\Delta \dot{x}_i = \sum_{j=1}^{n} a_{ij}\, \Delta x_j + \lambda_i\, l\,(t)\, x_1, \tag{5.129'}$$

where $l(t)$ is a periodic piecewise-constant function :

$$l(t) = K_a \quad (t_{a-1} < t < t_a\,, \quad a = 1,\ 2, \ldots,\ r).$$

We write the equation of the surfaces of discontinuity for equations (5.129′) in the form

$$x_1 - \sigma_a = 0.$$

* These limitations will not be mentioned here since they are always satisfied in the system (5.1) which we shall be interested in later.

Then the conditions (5.130) reduce to

$$\Delta x_i (t_a + 0) - \Delta x_i (t_a - 0) = R_i^a \Delta x_1 (t_a - 0), \qquad (5.130')$$

where

$$R_i^a = \varepsilon \lambda_i \xi \frac{1}{\dot{\tilde{x}} (t_a - 0)} .$$

Here ξ is the discontinuity in $f(x_1)$ for $x_1 = \tilde{x}_1 (t_a)$ and $\varepsilon = +1$ or -1 depending on the direction of intersection of the surface by the trajectory.

Integrating equations (5.129′) for $(t_{a-1} < t < t_a)$ and approvisioning them, taking (5.130′) into account at the ends of the period, we find linear relations expressing the values of the co-ordinate at the end of the period $\Delta x_i (t_0 + T)$ in terms of its value at the beginning of the period $\Delta x_i (t_0)$.

The solution $\Delta x_i = 0$ of the linear equation (5.129′) with periodic coefficients will be asymptotically stable (and, of course, so will the given periodic motion $\tilde{x}_1 (t)$ if for any small $\Delta x_1 (t_0)$ the condition

$$\Delta x_i (t_0 + T) < \Delta x_i (t_0).$$

is satisfied.

Thus, the question of the stability of the periodic solutions of an initial system (5.1) with a piecewise-linear characteristic can be solved by the approvision method if we apply it to the linear equation (5.129′) with piecewise-constant periodic coefficients, while taking the saltus conditions (5.130′) into account.

For those readers who are acquainted with matrix calculus we remark that if the result of this approvisioning can be written in matrix form

$$\Delta x (t_0 + T) = U \Delta x (t_0),$$

where U is a constant transformation matrix, it is necessary for the stability of the given periodic solution that the characteristic roots of the matrix U shall lie in the unit circle.

In the special case of the most simple symmetric state in a system with a symmetric relay characteristic (Section 9) the conditions for the stability of the periodic solution can be reduced to a simple criterion.

If the roots of the characteristic equation of the linear part of the system are known the characteristic equation answering the question of the stability of the periodic solutions is of the form

$$\sum_{j=1}^{n} C_j \frac{e^{p_j \frac{T}{2}}}{1 + e^{p_j \frac{T}{2}}} \frac{1}{\lambda + e^{p_j \frac{T}{2}}} = 0,$$

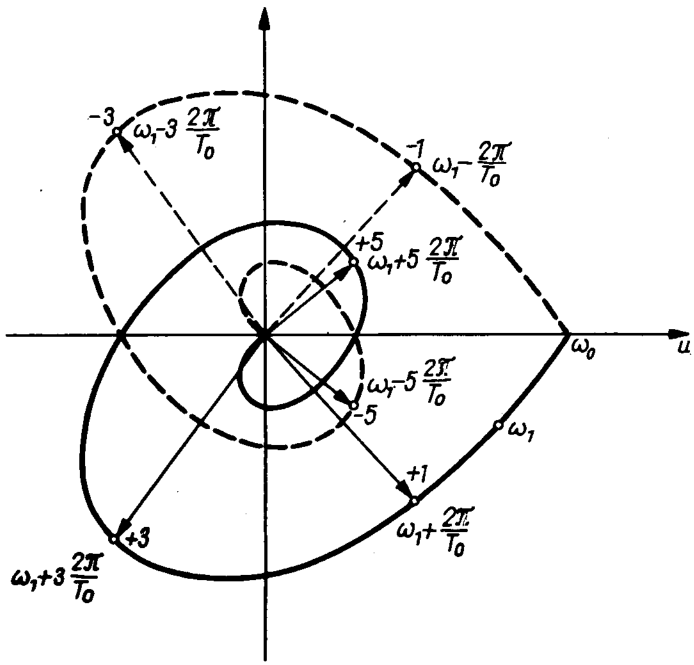


FIG. 225

where T is the period we have found, C_j are the coefficients, and p_j the roots of the characteristic equation of the linear part of the system $D(p) = 0$ and λ is unknown.

In order that all the periodic solutions shall be stable it is necessary and sufficient that in the λ-plane all the roots of this characteristic equation shall lie inside the circle of unit radius.

We can check that this condition is fulfilled by constructing the D-partition in the plane of one or two parameters. To do this, instead of putting $\lambda = i\,\omega$, we must carry out the substitution $\lambda = e^{i\omega}$, varying the values of ω from 0 to 2π.

If the roots are not known, but the amplitude-phase characteristic of the linear part of the system is known, in this case the question of the stability is answered by the construction of the hodograph

$$R(i\omega) = \sum_{m=-\infty}^{m=\infty} W\left[i\omega + i(2m-1)\frac{2\pi}{T}\right],$$

where

$$W(i\Omega) = \frac{K(i\Omega)}{D(i\Omega)}, \qquad \Omega = \omega + (2m-1)\frac{2\pi}{T}.$$

Setting some value $\omega = \omega_1$. finding on the hodograph $W(i\Omega)$ the points $\Omega = \omega_1 + \frac{2\pi}{T}$, $\Omega = \omega_1 + 3\frac{2\pi}{T}$, $\Omega = \omega_1 + 5\frac{2\pi}{T}$ and so on, and also $\Omega = \omega_1 - \frac{2\pi}{T}$, $\Omega = \omega_1 - 3\frac{2\pi}{T}$, $\Omega = \omega_1 - 5\frac{2\pi}{T}$ and so on (Fig. 225), and adding corresponding vectors, we find the vector $R(i\omega)$. Repeating this for another $\omega = \omega_2$ and so on we construct the hodograph $R(i\omega)$ for $0 \leqq \omega < \infty$.

The formulation of the criterion of stability of the periodic state is identical with that of the second amplitude criterion of stability* except that, instead of the hodograph $W(i\omega)$, we consider the hodograph $R(i\omega)$, and instead of the point -1 we consider the point r where

$$r = -\frac{\pi}{2k_p}\frac{T}{2\pi}\dot{x}\left(\frac{T}{2}\right).$$

It is only necessary to remember that for $\omega = 0$ the hodograph $W(i\omega)$ always passes through the point $-r$. The theory of stability shows that this fact is not an obstacle to stability.

The question of forced oscillations in relay systems is considered in an exactly similar way**.

12. Sliding Switchings in Systems with Piecewise-Linear Characteristics

In the previous sections, when applying the approvision method and the method which is based on looking for periodic solutions in the form of complete Fourier series, it was tacitly assumed that the

* See Chapter III.

** For more detail see Tsypkin, Y. Z. "The Theory of Relay Systems of Automatic Control", Gostekhizdat (1955).

crossing from one line of the characteristic to the other could always be made. This is not always the case, and it is now necessary to discuss this question in more detail.

Let us consider two contiguous branches of a characteristic (Fig. 226). First let $x_1 < \sigma_2$ and let the depicted point move along branch I. At the moment it reaches the point Q_1 with abscissa $x_1 = \sigma_1$ the depicted point crosses instantaneously to the branch II. Then the abscissa of the point P, where it lands on branch II after the jump, need not coincide with the abscissa of the point Q_1', i. e. it can differ

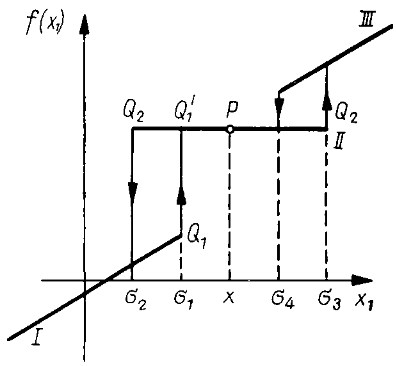


FIG. 226

from $x_1 = \sigma_1$. Also, the velocity, acceleration, etc. of the depicted point at the end of the jump also need not be the same, either in magnitude or in sign, as their values up to the time of the jump. All these quantities, the abscissa of the point P, the velocities of the depicted point after the jump, and so on, are determined from the saltus conditions (5.66).

Branch II of the characteristic can be given by a bounded segment. In general it need not contain any points satisfying the saltus conditions. The following six cases are therefore possible :

1. The point P exists on branch II, i.e. its abscissa x_P satisfied the inequality

$$\sigma_2 < x_P < \sigma_3 .$$

2. The point P coincides with Q_2 or with Q_2', but the velocity of the depicted point after the jump is directed "towards branch II", i.e.

$$x_P = \sigma_2 , \quad \dot{x}_P > 0 \quad \text{and} \quad x_P = \sigma_3 , \quad \dot{x}_P < 0 .$$

3. The point P coincides with Q_2, but the velocity of the depicted point after the jump is negative:

$$x_P = \sigma_2 , \qquad \dot{x}_P < 0 .$$

4. The point P coincides with Q_1', but its velocity is positive:

$$x_P = \sigma_3 , \qquad \dot{x}_P > 0 .$$

5. There is not in general any point P satisfying the saltus conditions and lying on the segment of branch II, but only on its continuation to the right, i.e. for

$$x_P > \sigma_3 .$$

6. There is not, in general, any point P on branch II, but one lies on its continuation to the left.

All six cases of the position of the point P and the sign of its velocity are represented in Fig. 227.

The crossing from one branch of the characteristic to the other is called *switching.*

In cases 1 and 2 the motion after the jump continues along branch II. Such switching is called *normal.*

In cases 3 and 6 immediately after the jump the depicted point lands at a point of descent of branch II or further to its left, under conditions when reverse crossing is performed from branch II to branch I. As a result of the jumps the depicted point, therefore, arrives not on branch II but once again on branch I. It slides backwards, so to speak, on to branch I. Such cases of switching are, therefore, called *sliding.*

In cases 4 and 5, the depicted point does not in general land on branch II, but also does not slide backwards on to branch I. Missing branch II it is thrown back immediately on to branch III if this exists. In the contrary case, x_1 is unbounded, and in theory increases instantaneously until the limits which always exists in real systems begin to be felt, and the equations of motion change. Switching of this kind is called *indeterminate*, since it does not even follow from the saltus conditions how the motion of the depicted point will continue afterwards.

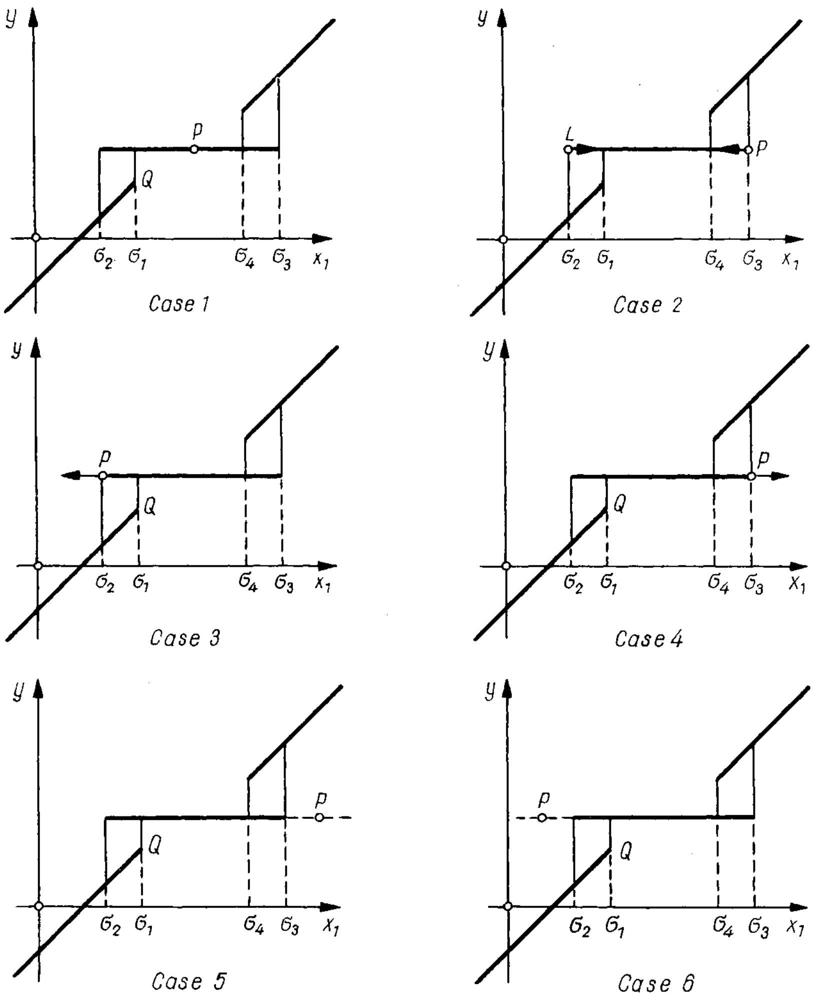


FIG. 227

In Sections 9 and 10 it was assumed that normal switching always takes place.

It is now necessary to determine the conditions under which switching will be normal.

We shall assume that the degrees n of the polynomial $D(p^*)$ and m of the polynomial $K(p^*)$ in the derived equation (5.86), can be

equal*. We consider the following three cases separately:

$$n - m > 1\,, \quad n - m = 1 \quad \text{and} \quad n - m = 0\,.$$

1. Suppose first that $n - m > 1$, i. e. $b_0 = b_1 = 0$. Then the first of the saltus conditions (5.66) is

$$a_0\,\xi_0 = b_0\,\eta_0 \tag{5.131}$$

and on putting $b_0 = 0$ we find immediately that $\xi_0 = 0$, i.e. during switching x_1 changes continuously and P has the abscissa $\sigma_P = \sigma_1$.

From the second saltus condition

$$a_0\,\xi_1 + a_1\,\xi_0 = b_0\,\eta_1 + b_1\,\eta_0\,, \tag{5.132}$$

and putting $b_0 = b_1 = \xi_0 = 0$ we obtain $\xi_1 = 0$. Consequently, in this case the velocity x_1 is continuous during switching. If the velocity $\dot{x}_1 > 0$ up to the beginning of switching, then it has this value also after switching (Fig. 228). *Thus, for* $n - m > 1$ *the switching is always normal.*

2. Now let $n - m = 1$, i.e. $b_0 = 0$, $b_1 \neq 0$. Then from (5.131) it follows as before that $\xi_0 = 0$, i.e. that $\sigma_P = \sigma_1$. From (5.132) we have

$$\xi_1 = \frac{b_1}{a_0}\,\eta_0,$$

where η_0 is the discontinuity in y on the characteristic at the point $x_1 = \sigma_1$ i.e. $\eta_0 = f(\sigma_1 + 0) - f(\sigma_1 - 0)$, and $\xi_1 = \dot{x}_1(\sigma_1 + 0) - \dot{x}_1\,(\sigma_1 - 0)$. Therefore, after switching $\dot{x}_1 > 0$ if

$$\frac{b_0}{a_0}\,\eta_0 > 0 \tag{5.133}$$

or if
$$\frac{b_0}{a_0}\,\eta_0 < 0, \text{ but} \left|\frac{b_0}{a_0}\right| \eta_0 < \dot{x}_1\,(\sigma_1 - 0). \tag{5.134}$$

After switching $\dot{x}_1 < 0$, if $\dfrac{b_0}{a_0}\,\eta_0 < 0$,

but
$$\left|\frac{b_0}{a_0}\right| \eta_0 > \dot{x}_1\,(\sigma_1 - 0). \tag{5.135}$$

* We shall explain below the conditions under which $n = m$.

If conditions (5.133) or (5.134) are satisfied switching is normal (Fig. 229). If the inequalities (5.135) are also satisfied the character of the switching is determined from the presence or absence of a loop in the characteristic. If there is no loop (i.e. $\sigma_2 = \sigma_1$) switching is sliding (Fig. 230). If there is a loop switching is normal (Fig. 231), but a peculiar motion then arises around the loop (Fig. 232) which changes into sliding switching in the limit as $\sigma_2 \rightarrow \sigma_1$.

3. Suppose, finally, that $n = m$, i.e. $b_0 \neq 0$. From (5.131) we obtain in this case

$$a_0 (x_P - \sigma_1) = b_0 (y_P - y_1), \tag{5.136}$$

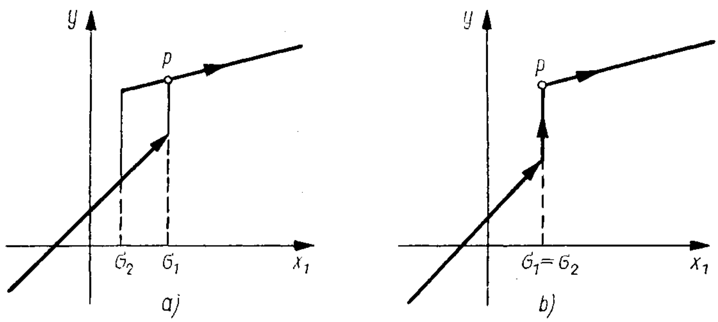


FIG. 228

where x_P, y_P are the coordinates of the point P (after the jump) and σ_1, $y_1 = f_1(\sigma_1)$ are the coordinates of the point Q_1 (before the jump.)

From (3.156) it follows that the point P must lie on the straight line

$$y_P = \frac{a_0}{b_0} x_P - \left(\frac{a_0}{b_0} \sigma_1 - y_1\right), \tag{5.137}$$

having a slope $\frac{a_0}{b_0}$ and passing through the point Q_1.

On the other hand, the point P must also lie on branch II of the characteristic. *Thus, the point P is determined by the intersection of branch II with the straight line (5.137).*

The switching is normal if branch II intersects the line (5.137), (Fig. 233a) ; it is sliding when the line (5.137) intersects the continuation of branch II to the left of $x_1 = \sigma_2$ (Fig. 233b) and is indeterminate if the line (5.137) intersects the continuation of branch II to the left of σ_3 (Fig. 233c).

Thus, for a relay characteristic (Fig. 234), for example, for $n = m$ the character of the switching is determined exclusively by the sign of $\frac{a_0}{b_0}$; it is normal when $\frac{a_0}{b_0} > 0$ (Fig. 234*d*), and sliding when $\frac{a_0}{b_0} < 0$ (Fig. 234*b*). For a relay characteristic there can be no indeterminate switching.

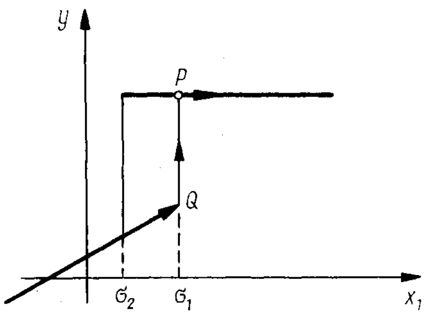


FIG. 229

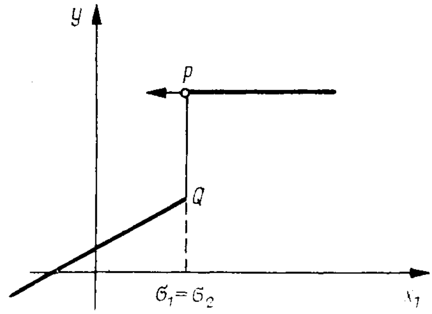


FIG. 230

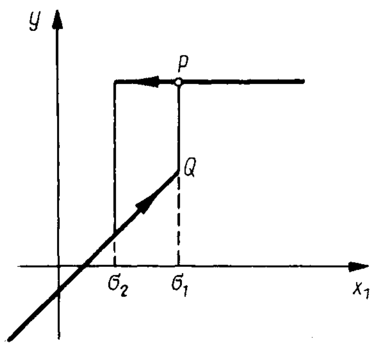


FIG. 231

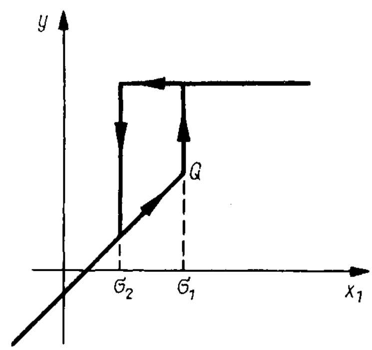


FIG. 232

We pause to investigate sliding switching in more detail. From what has been said it follows that two different types of sliding switching are possible. *Sliding switching of the first type* arises when $n - m = 1$ due to the change of the sign of $\dot{x}_1$ during switching. It is only possible in the absence of loops ($\sigma_1 = \sigma_2$) and is the limit of the motion arising "around the loop" as the width of the loop tends to zero ($\sigma_1 \rightarrow \sigma_2$). The sliding in this case occurs along a vertical straight line.

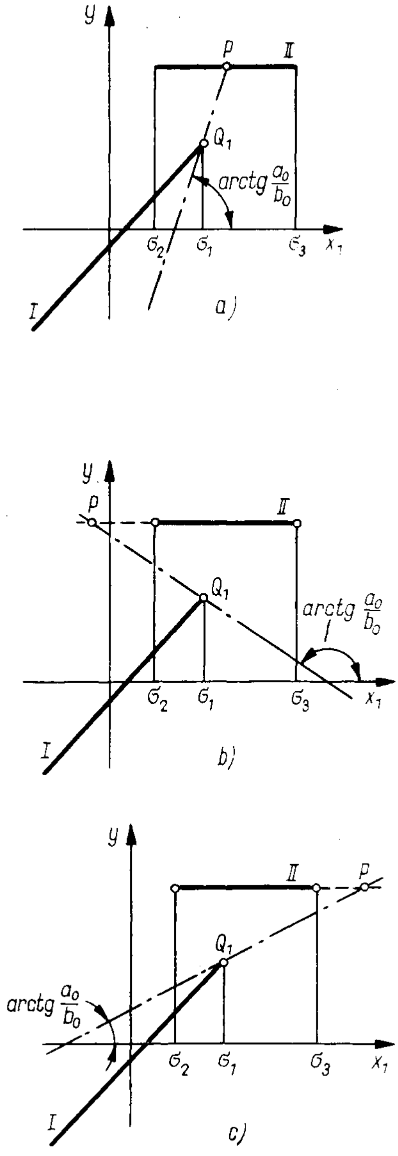


FIG. 233

Sliding switching of the second type arises for $n - m = 0$. It is completely independent of the change in the sign of the velocity $\dot{x}_1$ and is determined solely by the jump in x_1. It can occur also when

there are loops of any width; the presence of loops does not affect the character of the sliding motion and the width of the loop affects only the threshold value of $\frac{a_0}{b_0}$ at which sliding switching arises. Sliding switching of this type occurs along a line with slope $\frac{a_0}{b_0}$.

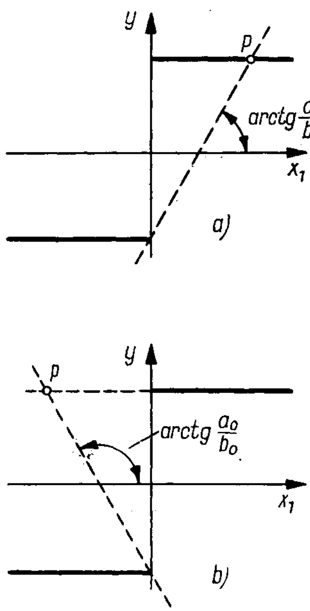


FIG. 234

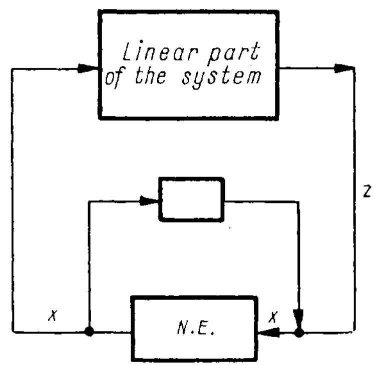


FIG. 235

From this account it follows that the case $n = m$ has a special significance for the switching conditions. But if the initial equations have the form (5.1) the equality $n = m$ is impossible. Let us suppose, however, that the non-linear element is closed by rigid feedback (Fig. 235), so that the equations of motion are

$$\left.\begin{aligned} D(P^*)z &= K(P^*)y, \\ y &= f(x), \\ x &= z - ry, \end{aligned}\right\} \qquad (5.138)$$

where $r =$ const.

These equations, when we eliminate z, reduce to

$$\left.\begin{aligned} D(P^*)\,x &= \overline{K(P^*)}\,y, \\ y &= f(x), \end{aligned}\right\} \tag{5.139}$$

where $\overline{K(P^*)} = K(P^*) - rD(P^*)$. Now in (5.139) the degree of $\overline{K(P^*)}$ is the same as the degree of $D(P^*)$.

If the feedback embracing the non-linear element is proportional, with transfer function $r = \dfrac{r_1}{\tau p + 1}$, then the equations of motion

$$\left.\begin{aligned} D(P^*)\,z &= K(P^*)\,y, \\ y &= f(x), \\ (Tp^* + 1)\,x &= (Tp^* + 1)\,z - ry \end{aligned}\right\}$$

reduce after z has been eliminated to the equations

$$\left.\begin{aligned} \overline{D(P^*)}\,x &= \overline{\overline{K(P^*)}}\,y, \\ y &= f(x), \end{aligned}\right\} \tag{5.140}$$

where

$$\overline{D(P^*)} = (Tp^* + 1)\,D(P^*), \text{ and}$$

$$\overline{\overline{K(P^*)}} = (Tp^* + 1)\,K(P^*) - r_1 D(P^*).$$

Now $n - m = 1$.

Thus, the case $n - m = 0$ always occurs when the non-linear element is closed by rigid feedback. The case $n - m = 1$ can also exist in the absence of feedback but it always exists if the non-linear element is closed by proportional feedback. *In systems containing feedback which closes the non-linear element it is necessary to follow the peculiarities of the crossing from one branch of the characteristic to the other with great care.*

In conclusion we note that the motion "around the loop" mentioned above* is sometimes used to improve the control process.

Thus for a relay characteristic, for example, with a zone of insensitivity and with loops, in such motion the operation of the system is accompanied by rapid switching on and off of a single contact of the relay and this often ensures a faster action and less overshoot.

* In the literature it sometimes is also called sliding.

13. Some Remarks Concerning the Phase Space of Dynamic Systems. The Value of Periodic Motions

What value has a knowledge of the periodic states in the elucidation of the "large" behaviour of the control system for real rather than for small disturbances?

Can we assert that in the absence of auto-oscillation states the system is "slightly" stable, stable, or "largely" stable?

To answer questions of this kind we must acquaint ourselves with the concept of the phase space and of phase trajectories of dynamic systems.

(a) *Phase portraits of linear systems*

We consider first the system which is described by the two differential equations of the first order:**

$$\left.\begin{aligned} \dot{x}_1 &= ax_1 + bx_2, \\ \dot{x}_2 &= cx_1 + dx_2, \end{aligned}\right\} \tag{5.141}$$

where a, b, c and d are given (and may be zero).

We assume that these equations have been integrated, i.e. that x_1 and x_2 have been found as function of the time t and of the initial conditions x_{10} and x_{20}:

$$\left.\begin{aligned} x_1 &= f_1(t,\ x_{10},\ x_{20}), \\ x_2 &= f_2(t,\ x_{10},\ x_{20}). \end{aligned}\right\} \tag{5.142}$$

For each fixed value of x_{10} and x_{20} equations (5.142) determine curves in the x_1, t-plane and in the x_2, t-plane (Fig. 236). For some other value of $\bar{x}_{10}$ and $\bar{x}_{20}$ these curves will have a different shape and may intersect the curves for x_{10} and x_{20} at various points.

The aggregate of equations (5.142) determines two two-parameter families of curves, in the x_1, t-plane and in the x_2, t-plane, an infinitely large number of curves passing through each point of the x_1-axis (or

** We recall in particular that the second order linear diferential equation $\ddot{x}_1 + h\dot{x}_1 + cx_1 = 0$ reduces to these if we put $\dot{x}_1 = x_2$, since in this case

$$\left.\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -cx_1 - hx_2 \end{aligned}\right\}$$

of the x_2-axis).* All these curves intersect one another. These graphs are not all suitable for representing the control process for various initial conditions.

It is considerably more convenient to use only one family of curves, eliminating t.

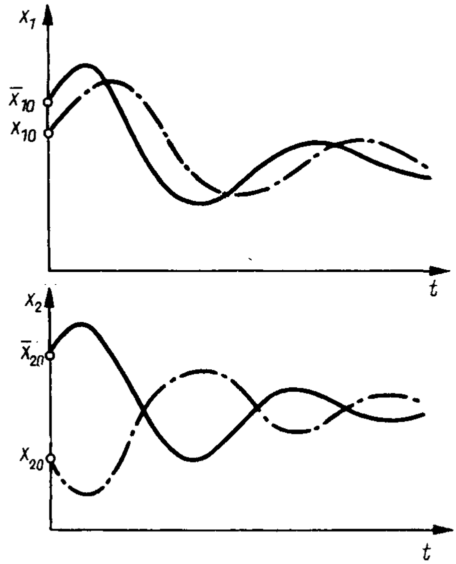


FIG. 236

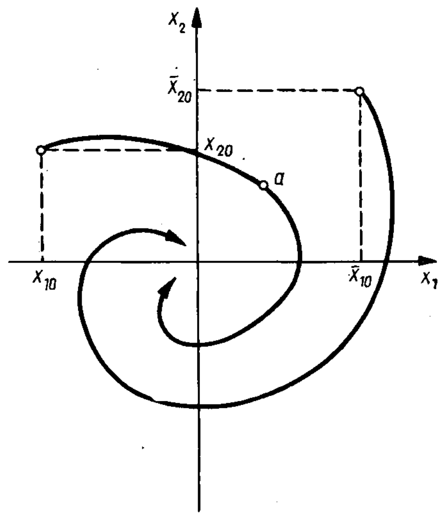


FIG. 237

We fix the values of x_{10} and x_{20} in equations (5.142) and consider these as parameter equations of curves in the x_1, x_2-plane (t being the parameter).

If we let t have any value, e. g. $t = t_1$, using equations (5.142) for fixed x_{10}, x_{20} we can calculate the values x_1 and x_2, i.e. a point in the x_1, x_2-plane. We give t another value $t = t_2$ say, and obtain a new x_1 and x_2, i.e. a new point in the x_1, x_2-plane. If we now vary t continuously from 0 to ∞ then the point (called the *depicted point)* will also move continuously in the x_1, x_2-plane and trace out a curve, the phase trajectory. If the system is stable, i.e. if as $t \to \infty$ both $x_1 \to 0$ and $x_2 \to 0$ then along the phase trajectory as $t \to \infty$ the depicted point tends to the coordinate origin (Fig. 237).

* Since for fixed x_{10} and variable x_{20} all the curves pass the same point of the x_1-axis.

In this plane we can construct in the same way the phase trajectory for any other values of $\bar{x}_{10}$ and $\bar{x}_{20}$. If the point corresponding to x_{10} and x_{20} lies on the earlier trajectory (for example, at the point a in Fig. 237), the depicted point will then move along this phase trajectory. If the point x_{10}, $\bar{x}_{20}$ turns out to be somewhere outside the constructed trajectory a new phase trajectory can be constructed by a similar method (Fig. 237).

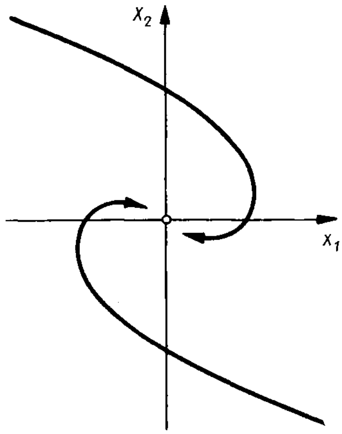


FIG. 238

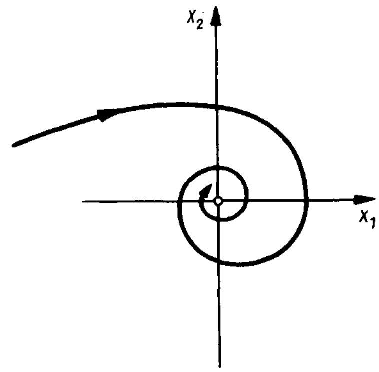


FIG. 239

Thus, for various x_{10} and x_{20} the equations (5.142) determine a family of curves which cover the x_1, x_2-plane everywhere densely. The x_1, x_2-plane is called the *phase plane* of the system. The phase plane, covered with the whole totality of phase trajectories, is called the *phase portrait* of the system.

In contrast to the family of integral curves (Fig. 236) the trajectories of the phase portrait (Fig. 237) can intersect only in a limited number of points, and only in one point, the origin of coordinates, for the linear case we have been considering up to now.

Indeed, the differential equation of the phase trajectories is obtained by dividing the first equation of the system (5.141) by the second:

$$\frac{dx_1}{dx_2} = \frac{ax_1 + cx_2}{cx_1 + dx_2}. \tag{5.143}$$

For any values of x_1 and x_2 (apart from $x_1 = x_2 = 0$) equation (5.143) determines a unique value of $\dfrac{dx_1}{dx_2}$, i.e. at any point of the

phase plane, apart from the origin, only one tangent can be drawn to the phase trajectory, and this proves that the phase trajectories do not intersect one another anywhere, except at the origin of coordinates.

At the origin of coordinates ($x_1 = x_2 = 0$) from equation (5.143) we obtain $\frac{dx_1}{dx_2} = \frac{0}{0}$, i.e. the origin is a singular point. It corresponds to the equilibrium of the system (at this point $\dot{x}_1 = \dot{x}_2 = 0$).

We recall that in a linear system the stability or instability does not depend on the magnitude of the initial deviations. On the strength

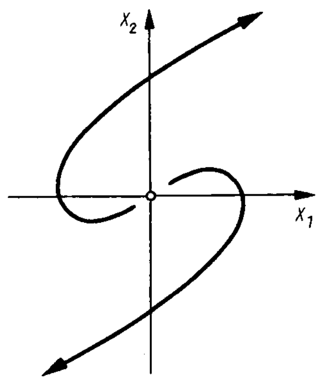


FIG. 240

of this, the phase portrait of a stable linear system is always such that the depicted point of any phase trajectory moves in the direction of the origin of coordinates. The "region of attraction" of the singular point of the origin of coordinates contains the whole phase plane. In unstable linear systems the depicted point at any point of the phase space on the phase trajectory recedes from the origin, and the "region of repulsion" of the singular point of the origin contains the whole phase plane.*

Figures 238 and 239 show two possible kinds of phase portrait for a stable linear system. For Fig. 238 the singular point is called the *stable node*, and for Fig. 239 the *stable focus*.

Figures 240, 241 and 242 show three possible kinds of phase portrait for an unstable linear system.

* With the exception of points lying on the line *aa* in the case of a saddle (see Fig. 242).

For Fig. 240 the singular point is called the *unstable node*, for Fig. 241 the *unstable focus*, and for Fig. 242 the *saddle point*.

Only one position of equilibrium is possible in a linear system, and correspondingly there is only one singular point, the origin of co-ordinates. Moreover, in a linear system undamped oscillations** are not possible, and so the phase portrait of the linear system does not contain closed phase trajectories.

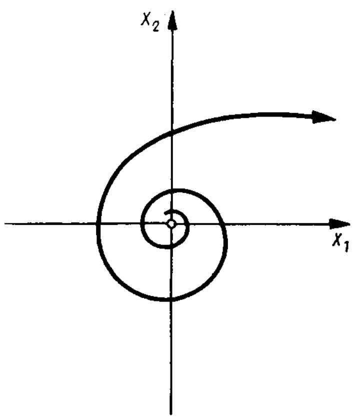


FIG. 241

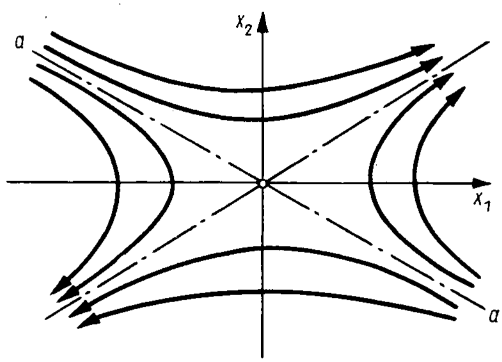


FIG. 242

For simplicity we have taken the system (5.141) consisting of two first order linear equations. Everything we have said can be extended at once to a system consisting of any number of linear first order differential equations and of course also to any linear system, since each higher order equation can be reduced to several first order equations by the use of the method given in the footnote on p. 210. It is only necessary to consider a phase space whose number of dimensions is equal to the number of first order equations in the considered system, instead of the two-dimensional phase plane.

It is not possible to represent the phase portrait of a system containing more than two equations in a plane which has only two dimensions, although the basic properties of the phase portraits described above for systems of two equations remain in force in this case also.

** We do not consider the system which corresponds exactly to the boundary of stability.

If the system is linear then:

(a) the phase space contains only one singular point, the origin of coordinates;

(b) if the system is stable, the "region of attraction" of this singular point (or if it is unstable, the "region of repulsion") contains the whole phase space;

(c) there are no closed phase trajectories in the phase space.

When the considered system of equations contains any one non-linear equation it is quite a different matter.

(b) *The phase portraits of a non-linear system*

Just as in a linear system, the control process described by equations containing non-linearities can be represented in a phase plane or in a phase space.

Here too we take as a basic example the case when the motion is described by two first order differential equations:

$$\left.\begin{aligned} \dot{x}_1 &= F_1(x_1, x_2), \\ \dot{x}_2 &= F_2(x_1, x_2), \end{aligned}\right\} \tag{5.144}$$

where $F_1(x_1, x_2)$ and $F_2(x_1, x_2)$ are in the general case) known non-linear functions of given arguments.

The differential equation of the phase trajectories is obtained by dividing the first equation of this system by the second:

$$\frac{\mathrm{d}x_1}{\mathrm{d}x_2} = \frac{F_1(x_1, x_2)}{F_2(x_1, x_2)}. \tag{5.145}$$

Only one tangent can be drawn to the phase trajectory, and therefore there are not necessarily points of intersection of the phase trajectories at all points of the phase plane where $F_1(x_1, x_2)$ and $F_2(x_1, x_2)$ are not simultaneously zero. The singular points of the system are found from the condition $\frac{\mathrm{d}x_1}{\mathrm{d}x_2} = \frac{0}{0}$, i. e. are the common roots of the two equations:

$$\left.\begin{aligned} F_1(x_1, x_2) &= 0, \\ F_2(x_1, x_2) &= 0. \end{aligned}\right\} \tag{5.146}$$

Previously, when we considered a linear system, we had

$$F_1(x_1, x_2) = ax_1 + bx_2\,,$$
$$F_2(x_1, x_2) = cx_1 + d\,x_2$$

and equations (5.146) had only one simultaneous solution: $x_1 = x_2 = 0$. In the x_1, x_2-plane the conditions (5.146) determine two straight lines intersecting at the origin (Fig. 243). But if the functions $F_1(x_1, x_2)$

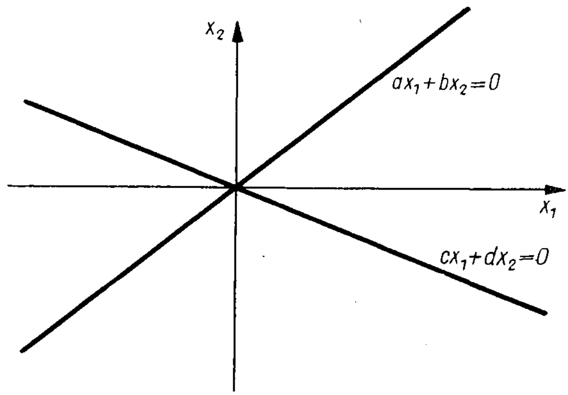


FIG. 243

and $F_2(x_1, x_2)$ are non-linear, then the curves given by (5.146) can intersect in a point different from the origin of coordinates as well. In this case equations (5.146) have solutions other than the solution* $x_1 = x_2 = 0$.

In this case positions of equilibrium different from the controlled state are also possible in the system (Fig. 244), and the character of the motion depends on the magnitude of the initial deviations.

An example is given in Fig. 245 of a phase portrait for the system (5.144) when the curves $F_1 = 0$ and $F_2 = 0$ intersect only in one point apart from the origin (which is a singular point of the "stable focus" type), this point being where the "saddle point" lies. The bold line shows the trajectory which passes through the "saddle point" and separates the "region of attraction" of the investigated stable equi-

* The curves $F_1 = 0$ and $F_2 = 0$ always intersect at the origin, since we have taken the position of equilibrium as that corresponding to $x_1 = x_2 = 0$, i.e. at this point $\dot{x}_1 = \dot{x}_2 = 0$.

librium state, i.e. of the singular point of the "stable focus" type lying at the origin of coordinates (this region is shaded in Fig. 245).

If the initial deviations in x_{10}, x_{20} give a point lying inside the shaded region in the phase plane (Fig. 245), after a time the depicted point on the phase trajectory will approach the origin of coordinates and the system will be stable with respect to this initial deviation. If, however, the initial deviation is such that the point x_{10}, x_{20} lies outside the shaded region the depicted point on its corresponding phase tra-

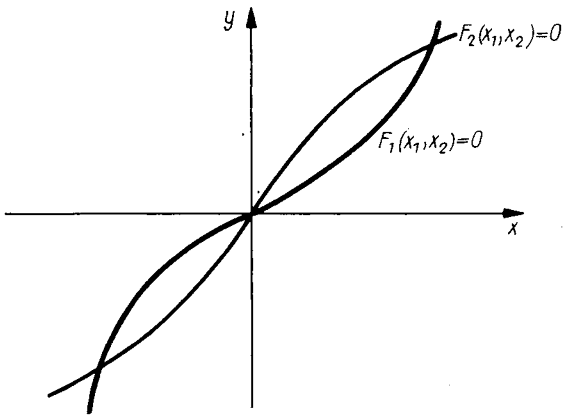


FIG. 244

jectory goes away to infinity and the system is unstable with respect to such an initial deviation. The region of attraction of the singular point at the origin can therefore be called the region of stability of the system.

The phase portrait shown in Fig. 245, just like the phase portrait of a linear system, does not contain closed phase trajectories. At the same time, in non-linear systems, both in the case when there is only one singular point and when there are several singular points, there can be closed trajectories.

Figure 246 gives an example of a system having only one singular point, at the origin of coordinates ("stable focus"), and one closed trajectory, surrounding the origin. The phase trajectories cannot intersect except at the singular point, and therefore the closed trajectory (we usually call it the limit cycle) separates the region of attraction or the region of stability of the singular point (shaded in Fig. 246). Inside the limit cycle the phase trajectories "wind together" with it

and "wind around" the origin of coordinates. Outside the phase trajectories they "unwind" from the limit cycle, and along any phase trajectory the depicted point goes off to infinity. The limit cycle itself corresponds to undamped oscillations, but in the case being considered they are unstable. However small the disturbance the depicted point, on leaving the limit cycle, will not return to it again,

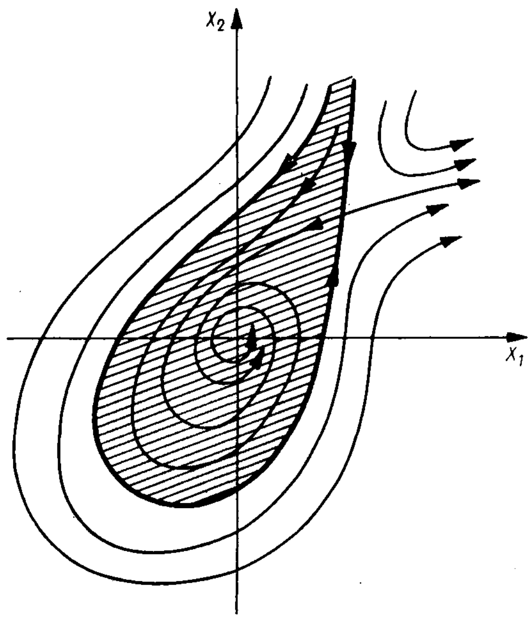


FIG. 245

and will move along the corresponding trajectory towards the origin of coordinates or to infinity. Undamped oscillations in such a system are never found in practice, and the limit cycle serves only to define the region of stability.

Another example of a system having one singular point ("unstable focus") and one limit cycle surrounding it, is shown in Fig. 247. In this case the investigated position of equilibrium is unstable, but the "region of repulsion" or the "region of instability" of the system is bounded by the limit cycle (this region is shaded in Fig. 247). After an initial deviation which lies inside the shaded region, the oscillations intensify and gradually undamped oscillations which correspond to

the limit cycle are set up. Conversely after initial deviations lying outside the shaded region, the oscillations die down until those corresponding to the limit cycle are restored. In this case the limit cycle

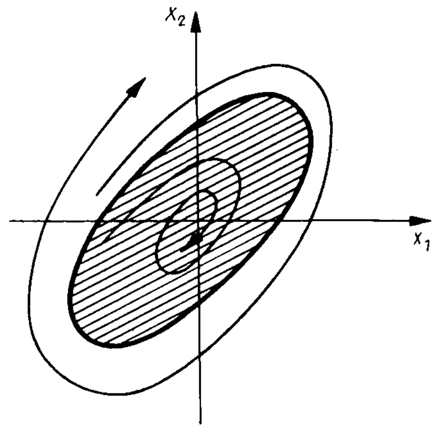


FIG. 246

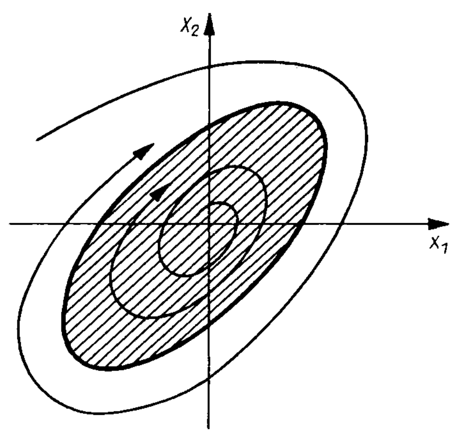


FIG. 247

not only isolates the region of instability, but also determines the stable undamped oscillations in the system which can be observed in practice.

Figure 248 gives an example of the phase portrait of a system containing two limit cycles surrounding the unique singular point of

the system ("stable focus"). In this case the region of stability of the singular point corresponding to the controlled equilibrium is separated by the inner unstable limit cycle. If in Fig. 248 the initial deviations in x_{10} and x_{20} determine a point within the inner limit cycle, the position of equilibrium is restored and the system is stable. If, however, this point is not inside the inner limit cycle, but lies outside it, undamped oscillations corresponding to the outer stable limit cycle are established in the system over the course of time.

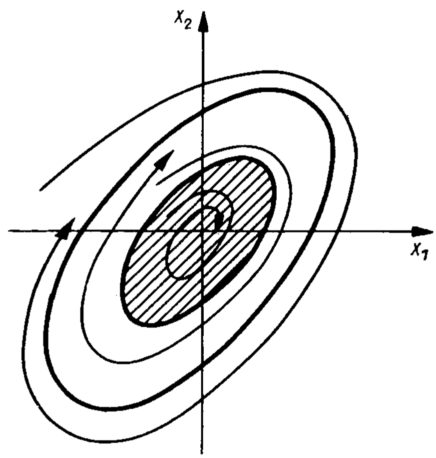


FIG. 248

Figure 249 shows a similar phase portrait for the case when the position of controlled equilibrium is unstable and a singular point of the "unstable focus" type lies at the origin of coordinates. In this case undamped oscillations corresponding to the inner limit cycle are set up after initial deviations x_{10} and x_{20} determining any point lying inside the outer limit cycle in the phase plane of Fig. 249. When the point lies outside, the amplitude of the oscillation grows without limit and undamped oscillations are not restored.

The system can also have limit cycles when the phase portrait contains more than one singular point. An example of this is shown in Fig. 250. In this case the controlled equilibrium is unstable (the origin is an "unstable" focus), and the limit cycle corresponds to stable undamped oscillations which are set up over the course of time if the initial deflections define a point inside the region separated by

the phase trajectory passing through the second singular point ("saddle"). This trajectory is shown by the bold line in Fig. 250.

If the phase portrait of the system contains more than one singular point or if it contains closed trajectories (limit cycles) then the region of stability cannot contain the whole phase plane as in the linear system. In this case the region of stability is always bounded by a limit cycle or phase trajectory passing through a singular point.

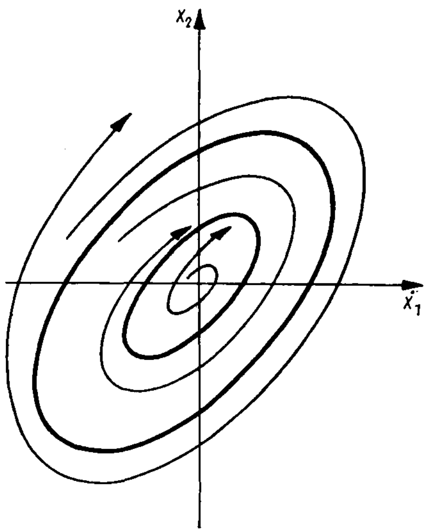


FIG. 249

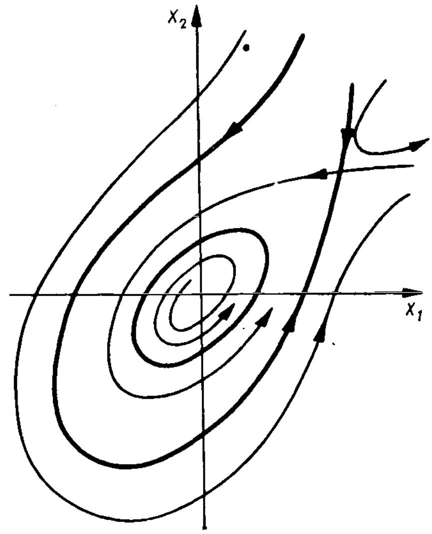


FIG. 250

Of course the phase portrait of a non-linear system (5.144) need not contain extra singular points or closed trajectories.

In this case the region of stability of the controlled equilibrium (the origin of coordinates of the phase space) can include the whole phase plane, and, just as in the linear case, whether there is stability need not depend on the magnitude of the initial deflection.

Up to now we have considered a system consisting of two first order equations.

In most practical problems of automatic control theory we have to deal with higher order equations.

If n, the order of the equation of the system, is larger than two, instead of a two-dimensional phase space we imagine an n-dimensional space, i. e. the space in which a point must be specified by n numbers.

If the given system of differential equations has a periodic solution a closed curve in the phase space corresponds to this solution.

In the plane the closed curves are the boundaries of regions. In the space surfaces and not curves must bound defined regions. As before a closed trajectory, therefore, corresponds to a periodic solution of the considered system of differential equations, but is not the boundary of a region.

There are two main differences between the phase plane and the phase space.

1. In the phase plane the limit cycle is not only a form of oscillatory motion, but also the boundary of the region of stability for another limit cycle or singular point.

Sometimes separatrix curves can be boundaries, but this happens comparatively rarely (mainly when there are several singular points, when the trajectories passing through the saddle point are separatrixes).

In the phase space no curve (not even the limit cycle) can be the boundary of the region.

The regions are bounded by separatrix surfaces, which entirely consist of phase trajectories.

Thus, in a phase plane the existence of singular points, limit cycles and the determination of their "slight" stability usually answer the question about their "large" stability too. In a phase space, to do this we must find also the separatrix surfaces, and this is a problem of extraordinary complexity

2. In systems of the second order the oscillations can only be periodic.

In higher order systems oscillations of different frequencies can exist, for example

$$x = A \sin \omega t + B \sin \Omega t\ .$$

If the frequencies ω and Ω are not connected by integer relations $n\,\omega + m\,\Omega = \mathrm{k}$ (where n, m and k are integers) then the sum of these two oscillations is also an oscillation, but is not periodic.

Such an oscillation in the phase space no longer determines a closed trajectory, but determines a trajectory which completely fills some closed volume (such as a torus). These trajectories can be stable or unstable.

(c) *"Slight", "large" and "unbounded" stability*

The concept of stability can be given a graphical interpretation.

The equilibrium is called "slightly" stable if a stable singular point in the phase space of the system corresponds to this equilibrium, i.e. if we can find a region in the phase space which is such that after any initial deflection belonging to it the depicted point approaches the singular point corresponding to the controlled equilibrium.

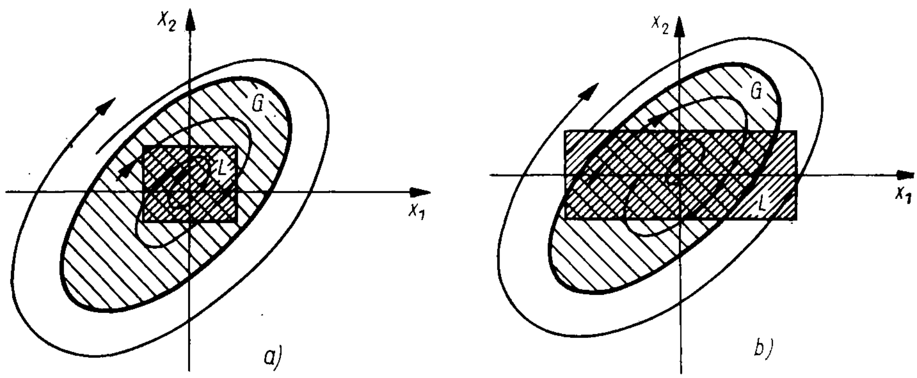


FIG. 251

Thus, by saying *that the controlled state is "slightly" stable, we are only asserting the presence of a stable singular point, and are not defining the boundaries of its attraction in any way.*

Let the phase portrait of the system be constructed, and the region of stability found. We call it the region G.

In the x_1, x_2-plane we shall now indicate a region of initial deflections which are possible in the considered system of automatic control with its given technical conditions. We call it the region L.

If all the points of L belong to G the controlled state is said to be *"largely" stable.*

Figure 251 gives an example of a phase portrait in which the region of stability is isolated by an unstable limit cycle, and the region of possible initial deflections is given in the form of a rectangle. In Fig. 251*b* the system is "slightly" stable but "largely" unstable, since there are several possible initial deflections for which the controlled state is not restored.

If the region of stability is unbounded and contains the whole phase space, i.e. if the system is stable after any initial deflection, then it is said to be unbounded stable.

In exactly the same we define "slight", "large" and "unbounded" stability of limit cycles.

(d) *The parameter space. Bifurcations*

Phase portraits are especially suitable for an estimate of the qualitative picture of the motions possible in the system. One glance at Fig. 248 for example is sufficient for us to assert that in the system with this phase portrait the position of equilibrium is stable with respect to initial deviations which do not go outside the defined threshold, but that after initial deflections outside this threshold undamped oscillations are set up, their amplitude and frequency being completely independent of the amount by which the given threshold was exceeded. This picture is used whatever the size of the limit cycles (only the threshold values of the initial deviations, and the amplitude and frequency of the undamped oscillations, depend on their size, i.e. the quantitative and not the qualitative aspect).

The qualitative picture of the motion as a whole is given by the *topological structure of the phase space*, i.e. by the presence, type and mutual distribution of the singular trajectories, the singular points, limit cycles, and separatrixes.

Let us now fix the values of all the parameters except one (the time constant or the coefficient of amplification of any stage, for example). Let us call this parameter a. On the numerical axis of a we select the point $a = a_1$. For this value $a = a_1$ all the parameters of the system are defined and by one method or another the phase portrait of the system can be constructed. Suppose it is of the form shown in Fig. 248, for example. We now change the value of a. To this new value of a there corresponds a new phase portrait. Thus, for each point of the numerical axis of a a corresponding defined phase portrait can be represented. From the theorem on the continuous dependence of integrals of differential equations on their parameters, it follows that for a small change in a the phase portrait also changes little. If, for example, the phase portrait having the topological structure shown in Fig. 248 corresponds to $a = a_1$, the phase portrait for

$a = a_1 + \varepsilon$ will have the same portrait, at least if ε is sufficiently small. As $|\varepsilon|$ increases we reach a value of ε for which the structure of the phase portrait changes. This can happen, for example, when the separate singular trajectories fuse, or when their stability changes. Thus in the phase portrait of Fig. 248, for example, the inner limit cycle can, as a changes, contract towards the origin of coordinates, and for some value $a = a^*$ fuses with it, so that for a further change in a the phase portrait will contain only one limit cycle, and the singular point at the origin of co-ordinates become unstable. This new structure of the phase space will be retained for larger values of a if no new value $a = a^{**}$ is reached for which the topological structure of the phase portrait changes again.

Thus when any parameter changes the quantitative characteristics of the phase portrait change continuously, but the qualitative characteristics of the phase portrait, its topological structure, undergo sharp changes for discrete values of the parameter. These discrete values of the parameter are called *bifurcations*.

Points of bifurcation divide the numerical axis of the parameter a into sections corresponding to systems having identical structures in the phase space.

In linear systems there are only two possible topological structures in the phase space: there are no singular trajectories, the only singular point is at the origin (stable for one structure and unstable for the other), and the region of stability or instability is not bounded. A change in the topological structure occurs for discrete values of the varied parameter on the boundary of the region of stability. In this sense the value of the parameter on the boundary of the region of stability is the bifurcation. But in a non-linear system, the concept of bifurcation is more general. The bifurcation value of the parameters can correspond not only to the change of stability of the singular point, but also to the vanishing or birth of a limit cycle, to the change in the number of singular points, and so on.

Up to now, for simplicity, we have discussed the numerical axis, i.e. the space of a single parameter. There is no difficulty in generalizing the concept to parameter spaces of any number of dimensions, just as we generalized the concept of the region of stability in linear systems.

Suppose that m parameters must be given for the complete determination of the equations of motion of the system.

In the m-dimensional space of these parameters each point corresponds to a definite structure of the phase space and, consequently, a hyper-surface dividing the parameter space into regions corresponding to systems whose phase spaces have identical topological structure can be found. In the special case when the considered system can be uniquely determined by two given parameters, the space of the parameters is the usual plane, and the bifurcation values of the parameters determine a curve in this plane.

(e) *The concept of complete and special solutions of non-linear problems. The value of periodic volutions*

The concept of the "complete solution of a non-linear problem of automatic control theory" is conventional. The more perfect the methods of the theory of automatic control are, the more complete will be the information about the dynamic system.

At the present time a non-linear problem is considered qualitatively as being completely solved if the possible phase portraits have been determined, and bifurcation boundaries have been defined in the parameter space. The quantitative solution of the problem requires, in addition, the determination and positioning of the limit cycles and separatrixes (or separatrix surfaces) for each point of the parameter space.

So complete a solution of a non-linear problem has only been found for individual special cases and, as a rule, when the problem is idealized. Because of this we are often satisfied by the solution of particular problems. Two particular problems have acquired great value:

(a) The problem of determining the conditions for which the phase portrait of the system does not contain any peculiarities, apart from the singular point corresponding to the controlled equilibrium.*

(b) The problem of determining the periodic solutions of the differential equations describing the control process, the conditions of their existence and of their "slight" stability, i.e. the problem of the existence of the limit cycles and those parts of the bifurcation boundaries in the parameter space which correspond to a change in the number of cycles.

* This problem was mentioned at the end of Chapter III.

We can now return to the questions formulated at the beginning of this section, concerning the worth of the determination of the periodic solutions and of their stability.

In any general case, the determination of all the singular points and periodic solutions and the conditions for their local stability is insufficient for us to find all the possible types of motion in the considered system. Only in the special case of second order systems, when the phase plane does not contain separatrix curves, do the value of the singular point, the periodic solutions and the conditions of their existence and stability, enable us to solve the problem completely, i.e. to determine the possible topological structures of the phase portraits and the partition of the parameter space by the bifurcation boundaries into regions corresponding to identical phase portraits.

Only in these cases does the absence of periodic solutions testify that the system is "slightly" stable, stable, or "largely" stable. In other cases, when the order of the system of equations is higher than two, the fact that the system does not have periodic solutions, but that the equilibrium is stable, still does not necessarily mean that the system is also "largely" stable.

14. Concluding Remarks

A linear analysis enables us to evaluate the stability of the system with respect to small disturbances and to investigate the character of the control process, provided the disturbing actions are sufficiently small. Usually the investigation of the control process for real disturbances requires us to take non-linearities into account, i.e. to consider non-linear differential equations. The problem of constructing the process in non-linear systems with given conditions is solved by graphical and numerical methods, which are explained in detail in any course on approximate and numerical analysis. General methods of synthesis and analysis of non-linear systems of control have as yet hardly been developed, and have still not been sufficiently checked in practice for there to be any point in including them in this short course on automatic control. But one of the special questions arising in the consideration of non-linear control systems, the question of the steady periodic states, has been developed in very great detail and has led to methods suitable for technical calculations.

In the absence of external periodic actions the periodic states (auto-oscillations) are usually harmful, and the problem consists in determining the values of the parameters for which the auto-oscillations do not arise. In several cases, though, the auto-oscillations are useful or sometimes harmless, because the principle of action itself of the device is based on the use of auto-oscillations (in vibratory controllers, two-position controllers, etc.). In such cases the aim of the investigation consists in finding the parameters of the auto-oscillations (the amplitude and frequency) and in discovering how they can be changed in the required direction. When an external periodic action acts on the system periodic states arise only for frequencies lying within a defined strip; we then have to find the amplitude and phase of the periodic response of the system (of the "forced oscillations") and the boundaries of this strip.

The methods used to solve these problems can be subdivided into two large groups. To the first group belong approximate methods based on the assumption that the investigated periodic state is nearly harmonic. To the second group belong exact methods, which do not ignore the harmonics in the Fourier expansion of the periodic solution.

An approximate method can be used only if the assumption that the investigated state is nearly harmonic can be justified. Such an assumption can be made in two cases: when the system is little different from a linear system in which the amplitude characteristic has a large and sharp peak ("auto-resonance") and when the linear part of the given system blocks the harmonics caused by the non-linear element ("filter").

When auto-resonance justifies the assumption about the nearly harmonic nature of the oscillations, we can consider that the setting-up process represents a special "almost-harmonic" process, i.e. sinusoidal with slowly changing amplitude and phase. This enables us to obtain simple criteria for the stability of the periodic states. These criteria are not suitable when a filter causes the oscillations to be nearly harmonic, since with a filter it is not possible to consider the setting-up process as "almost-harmonic". The approximate method leads to the following main deduction. When the non-linear characteristic is odd, the frequency of the auto-oscillations does not depend on its shape; the shape of the non-linear characteristic affects only the amplitude of the auto-oscillations.

The periodic states can be found exactly without ignoring the harmonics when the non-linear characteristic consists of straight line segments. To do this we can use the approvision method, or we can look for periodic solutions in the form of complete Fourier series (without ignoring any harmonics). Both in the first and the second case the problem reduces to the formation of the period equations, which are systems of transcendental equations in terms of the times of occurrence of the separate segments of the characteristic during the periodic state. Only in one case, in the investigation of the simplest symmetric periodic state in a system with a symmetric relay characteristic, can we look for only one time, the period of the oscillations. In this case alone, instead of the system of period equations we therefore obtain a single period equation and we can solve it graphically relatively simply. In the other cases the period equations are solved on machines, or by tedious numerical and graphical methods. The problem of the stability of the periodic solutions which have been found exactly is solved by the equations of linear approximation. To do this we form linear equations with piecewise-constant coefficients, together with linear relations between the discontinuities. This system of linear equations is integrated by the approvision method over the limits of one period. The integral found by this method enables us to judge the stability of the investigated periodic solution.

The existence or absence of periodic states is not directly connected with the determination of the boundaries of the region of stability of the controlled state when the order of the system of equations describing the process is higher than two. The region of stability can be bounded, and the system can be "largely" unstable, in spite of the fact that it is "slightly" stable and that there are no periodic states. The sufficient conditions for "large" stability can be obtained from other expressions, described at the end of Chapter III.

APPENDIX 1

LAPLACE AND FOURIER TRANSFORMS AND THEIR APPLICATION TO THE INTEGRATION OF SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS

1. A General Introduction to the Laplace Transform

Let the function $f(t)$ be given, equal to zero for $t < 0$ and different from zero for all, or at least some, values of $t \geqslant 0$. Then the Laplace integral of the function $f(t)$ is the integral of the form

$$\int_0^\infty f(t)\, e^{-pt}\, dt,$$

where p is a complex number.

We denote this integral by $L\,[f(t)]$.

If $f(t)$ is given we can calculate the value of $L\,[f(t)]$ for each value of the number p. In this sense, $L\,[f(t)]$ is a function of p. A definite function $L\,[f(t)]$ corresponds to each $f(t)$, with some general restrictions laid upon it.*

$f(t)$ is called the *original* function, and $L\,[f(t)]$ is called the *Laplace transform or representation* of the function We write this

$$L\,[f(t)] \doteqdot f(t).$$

The determination of the transform of the original function is the direct problem of the theory of the Laplace transform. The converse

* The restrictions which are put on the function $f(t)$ in order that its transform exists and, conversely, which are put on the transform in order that the function shall exist, are not discussed here. In control problems these restrictions are not important. For more detail see the bibliography to the Appendix.

problem is the determination of the original function from the transform.

In the theory of automatic control the most important use of the Laplace transform is in simplifying the integration of systems of linear differential equations with constant coefficients.

There are several methods by which, knowing the transform of one function, we can find that of another function. Only some of these methods are important for what follows. We list them.*

1. *The determination of the transform of the sum of several functions from their separate transforms*

If

$$L[f_1(t)] \doteqdot f_1(t);\quad L[f_2(t)] \doteqdot f_2(t);\ldots; L[(f_n(t)] \doteqdot f_n(t),$$

then

$$L[f_1(t) + f_2(t) + \cdots + f_n(t)] = L[f_1(t)] + L[f_2(t)] + \cdots + L[f_n(t)].$$

2. *The determination of the transform of the derivative of a given function*

Let

$$L[f(t)] \doteqdot f(t).$$

Then

$$L\left[\frac{\mathrm{d}f(t)}{\mathrm{d}t}\right] \doteqdot pL[f(t)] - f(0).$$

If $f(0) = 0$, then $L\left[\frac{\mathrm{d}f(t)}{\mathrm{d}t}\right] \doteqdot pL[f(t)]$. Similarly

$$L\left[\frac{\mathrm{d}^2 f(t)}{\mathrm{d}t^2}\right] \doteqdot p^2 L[f(t)] - \left[pf(0) + \frac{\mathrm{d}f(0)}{\mathrm{d}t}\right].$$

If*

$$f(0) = \frac{\mathrm{d}f(0)}{\mathrm{d}t} = 0,$$

* For their proofs see the bibliography to Appendix 1.

* In Appendix 1 we shall everywhere for short put $\frac{\mathrm{d}^r f(0)}{\mathrm{d}t^r}$ where $r = 1, 2, \ldots$ in place of $\left[\frac{\mathrm{d}^r f(t)}{\mathrm{d}t^r}\right]_{t=0}$

then

$$L\left[\frac{d^2 f(t)}{dt^2}\right] \div p^2 L[f(t)].$$

In general

$$L\left[\frac{d^m f(t)}{dt^m}\right] = p^m L[f(t)] -$$

$$-\left[p^{m-1} f(0) + p^{m-2}\frac{df(0)}{dt} + p^{m-3}\frac{d^2 f(0)}{dt^2} + \ldots + \frac{d^{m-1}f(0)}{dt^{m-1}}\right].$$

If

$$f(0) = \frac{df(0)}{dt} = \ldots = \frac{d^{m-1}f(0)}{dt^{m-1}} = 0,$$

then

$$L\left[\frac{d^m f(t)}{dt^m}\right] = p^m L[f(t)].$$

3. *The transform of a definite integral*

If

$$L[f(t)] \div f(t),$$

then

$$L\left[\int_0^\infty f(t)\,dt\right] \div \frac{L[f(t)]}{p}.$$

4. *The transforms of the basic functions encountered in control problems*

In Table A. 1 we set out the Laplace transforms of some functions which are encountered in control theory.

These expressions can easily be verified by substituting the corresponding $f(t)$ in the integral which defines the Laplace transform and integrating directly.

In the above table *1* denotes a function of t which is equal to zero for $t < 0$ and equal to 1 for $t \geqslant 0$ ("the unit function").

TABLE A. 1

Original	Transform
$L[\text{const}\,\mathbf{1}]$	$\dfrac{\text{const}}{p}$
$L[t^n\,\mathbf{1}]$	$\dfrac{n!}{p^{n+1}}$
$L[\mathbf{1}\sin\omega t]$	$\dfrac{\omega}{p^2+\omega^2}$
$L[\mathbf{1}\cos\omega t]$	$\dfrac{p}{p^2+\omega^2}$
$L[\mathbf{1}\sin(\omega t\pm\varphi)]$	$\dfrac{\omega\cos\varphi\pm p\sin\varphi}{p^2+\omega^2}$
$L[\mathbf{1}\cos(\omega t\pm\varphi)]$	$\dfrac{p\cos\varphi\mp\omega\sin\varphi}{p^2+\omega^2}$
$L\left[\mathbf{1}\dfrac{1}{a}\sinh at\right]=L\left[\mathbf{1}\dfrac{1}{2a}(e^{at}-e^{-at})\right]$	$\dfrac{1}{p^2-a^2}$
$L[\mathbf{1}\cosh at]=L\left[\mathbf{1}\dfrac{1}{2}(e^{at}+e^{-at})\right]$	$\dfrac{p}{p^2+a^2}$
$L[\mathbf{1}\,e^{at}]$	$\dfrac{1}{p-a}$
$L\left\{\mathbf{1}e^{-\frac{b}{2}t}\left[M\cos\sqrt{c-\dfrac{b^2}{4}}\,t+\dfrac{N-\dfrac{1}{2}Mb}{\sqrt{c-\dfrac{b^2}{4}}}\sin\sqrt{c-\dfrac{b^2}{4}}\,t\right]\right\}$	$\dfrac{Mp+N}{p^2+bp+c}$

5. *The construction of the transform of a function with a delayed argument*

If

$$L[f(t)] \doteq f(t),$$

then

$$e^{-\tau p} L[f(t)] \doteq f(t-\tau),$$

where

$$\tau = \text{const}.$$

6. *The construction of the original function of the product of two transformed functions*

If

$$L[f_1(t)] \doteq f_1(t) \text{ and } L[f_2(t)] \doteq f_2(t).$$

then

$$L[f_1(t)]\, L[f_2(t)] \doteq \int_0^t f_1(t-\tau) f_2(\tau)\, d\tau = \int_0^t f_1(\tau) f_2(t-\tau)\, d\tau.$$

An integral of this kind is said to be the convolution of the two functions.

To the product of two transforms corresponds not to the product of the original functions but their convolution.

7. *The limit properties of Laplace transforms*

Let $F(p) = L[f(t) \doteq f(t)$ and the real part of all the poles* of $F(p)$ be negative. Then

$$[f(t)]_{t=0} = [pF(p)]_{p=\infty}.$$

and, conversely,

$$[f(t)]_{t=\infty} = [pF(p)]_{p=0},$$

i.e. the limiting values of the function $f(t)$ can be found by putting the values $p = 0$ or $p = \infty$ in its transform and multiplying it by p. We can use these formulae to determine the limiting values of the solution of a differential equation from the form of this equation, without actually solving it.

* i.e., when $F(p)$ is a rational, fractional function, all the roots of its denominator.

2. The Integration of a Single Differential Equation

We consider a differential equation which has a right-hand side:

$$a_0 \frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1} x(t)}{\mathrm{d}t^{n-1}} + a_2 \frac{\mathrm{d}^{n-2} x(t)}{\mathrm{d}t^{n-2}} + \ldots + a_n x(t) = f(t). \tag{A.1}$$

We multiply both sides of the equation by e^{-pt} and integrate it from 0 to ∞:

$$\int_0^\infty \left[a_0 \frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1} x(t)}{\mathrm{d}t^{n-1}} + \ldots + a_n x(t)\right] e^{-pt}\mathrm{d}t = \int_0^\infty f(t)\, e^{-pt}\, dt,$$

or

$$a_0 \int_0^\infty \frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n} e^{-pt}\, \mathrm{d}t + a_1 \int_0^\infty \frac{\mathrm{d}^{n-1} x(t)}{\mathrm{d}t^{n-1}} e^{-pt}\, \mathrm{d}t + \ldots$$

$$\ldots + a_n \int_0^\infty x(t)\, e^{-pt}\, \mathrm{d}t = \int_0^\infty f(t)\, e^{-pt}\, \mathrm{d}t.$$

We can now write this equation in the form

$$a_0 L\left[\frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}\right] + a_1 L\left[\frac{\mathrm{d}^{n-1} x(t)}{\mathrm{d}t^{n-1}}\right] + \ldots + a_n L[x(t)] = L[f(t)].$$

Using the formulae given above we now replace the transforms of the derivatives by transforms of the primitive functions and the corresponding initial conditions:

$$a_0 p^n L[x(t)] - a_0\left[p^{n-1} x(0) + p^{n-2} \frac{\mathrm{d}x(0)}{\mathrm{d}t} + \ldots\right] +$$

$$+ a_1 p^{n-1} L[x(t)] - a_1\left[p^{n-2} x(0) + p^{n-3} \frac{dx(0)}{dt} + \ldots\right] +$$

$$+ \ldots + a_n L[x(t)] = L[f(t)].$$

If we now take all the terms involving the initial conditions to the right-hand side, and then take $L\,[x(t)]$ outside the brackets on the left-hand side, we obtain:

$$(a_0\,p^n + a_1\,p^{n-1} + \ldots + a_n)\,L\,[x\,(t)] = L\,[f\,(t)] + R\,(p)\,. \tag{A.2}$$

$R(p)$ here denotes the sum of all the terms involving initial conditions, and is a polynomial in p with coefficients which depend on the initial conditions.

If we collect like terms in $R(p)$ we obtain a polynomial in p:

$$R\,(p) = b_1\,p^{n-1} + \ldots + b_{n-1}\,p + b_n\,, \tag{A.3}$$

whose coefficients depend on the initial conditions and are given by

$$\left.\begin{aligned}
b_1 &= a_0\,x\,(0)\,,\\
b_2 &= a_0\,\frac{\mathrm{d}x\,(0)}{\mathrm{d}t} + a_1\,x\,(0)\,,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
b_{n-1} &= a_0\,\frac{\mathrm{d}^{n-2}\,x\,(0)}{\mathrm{d}t^{n-2}} + a_1\,\frac{\mathrm{d}^{n-3}\,x\,(0)}{\mathrm{d}t^{n-3}} + \ldots + a_{n-2}\,x\,(0)\,,\\
b_n &= a_0\,\frac{\mathrm{d}^{n-1}\,x\,(0)}{\mathrm{d}t^{n-1}} + a_1\,\frac{\mathrm{d}^{n-2}\,x\,(0)}{\mathrm{d}t^{n-2}} + \ldots + a_{n-1}\,x\,(0)\,.
\end{aligned}\right\} \tag{A.4}$$

If all the initial values are zero, i. e. if

$$x\,(0) = \frac{\mathrm{d}x\,(0)}{\mathrm{d}t} = \frac{\mathrm{d}^2\,x\,(0)}{\mathrm{d}t^2} = \ldots = \frac{\mathrm{d}^{n-1}\,x\,(0)}{\mathrm{d}t^{n-1}} = 0\,.$$

then $R(p)$ is identically zero, $R(p) = 0$.

The polynomial in the brackets on the left-hand side of equation (A.2) can be obtained directly from the given equation (A.1) if the mth derivative of x is denoted by $p^m x$ instead of by the usual $\dfrac{\mathrm{d}^m\,x}{\mathrm{d}t^m}$. Then equation (A.1) becomes

$$a_0\,p^n\,x + a_1\,p^{n-1}\,x + \ldots + a_n\,x = f\,(t)$$

or, taking x outside the brackets,

$$D(p)x = f(t)\,,$$

where

$$D\,(p) = a_0\,p^n + a_1\,p^{n-1} + \ldots + a_n.$$

The equation $D(p) = 0$ can be called the *characteristic* equation if p is regarded as a normal variable.

Equation (A.2) can now be written

$$D(p)\,L\,[x(t)] = L\,[f(t)] + R(p)$$

or

$$L\,[x\,(t)] = \frac{L\,[f\,(t)]}{D\,(p)} + \frac{R\,(p)}{D\,(p)}\,. \tag{A.5}$$

The next problem is to determine $x(t)$ from this transform. The term $\dfrac{L\,(f\,(t)]}{D\,(p)}$ determines the motion of the system under the action of a disturbance $f(t) \not\equiv 0$ with zero initial conditions, and the term $\dfrac{R\,(p)}{D\,(p)}$ determines the motion of the system which is stipulated by the initial conditions not being zero.

EXAMPLE 1. To find the transform of the integral of

$$a_0\,\frac{\mathrm{d}^2\,x\,(t)}{\mathrm{d}t^2} + a_1\,\frac{\mathrm{d}x\,(t)}{\mathrm{d}t} + a_2\,x\,(t) = 5\sin t\,,$$

when, for $t = 0$ the initial conditions are equal to

$$x = x_0\,,\ \ \frac{\mathrm{d}x}{\mathrm{d}t} = 0\,.$$

Multiplying both sides of the equation by e^{-pt} and integrating it between 0 and ∞, we obtain:

$$a_0\,L\left[\frac{\mathrm{d}^2\,x\,(t)}{\mathrm{d}t^2}\right] + a_1\,L\left[\frac{\mathrm{d}x\,(t)}{\mathrm{d}t}\right] + a_2\,L\,[x\,(t)] = 5L\,[\sin t]\,.$$

Performing the transformation, we find:

$$a_0 \left\{ p^2 L[x(t)] - \left[px(0) + \frac{\mathrm{d}x(0)}{\mathrm{d}t} \right] \right\}$$
$$+ a_1 \left\{ pL[x(t)] - x(0) \right\} + a_2 L[x(t)] = 5 \frac{1}{p^2 + 1},$$

or

$$(a_0 p^2 + a_1 p + a_2) L[x(t)] = \frac{5}{p^2 + 1} +$$
$$+ \left[a_0 px(0) + a_1 x(0) + a_0 \frac{\mathrm{d}x(0)}{\mathrm{d}t} \right].$$

Solving this equation for $L[x(t)]$, we obtain:

$$L[x(t)] = \frac{\dfrac{5}{p^2 + 1} + R(p)}{a_0 p^2 + a_1 p + a_2},$$

where

$$R(p) = a_0 px(0) + a_1 x(0) + a_0 \frac{\mathrm{d}x(0)}{\mathrm{d}t},$$

It is then necessary to go from the obtained transform to the original. To study a method for doing this in this case we return to equation (A.5), which defines the transform of the required function $x(t)$:

$$L[x(t)] = \frac{L[f(t)]}{D(p)} + \frac{R(p)}{D(p)},$$

and we put

$$f(t) \equiv 0.$$

Then

$$L[x(t)] = \frac{R(p)}{D(p)}.$$

We find a function $x(t)$ for which

$$L[x(t)] = \frac{R(p)}{D(p)} = \frac{b_1 p^{n-1} + b_2 p^{n-2} + \ldots + b_n}{a_0 p^n + a_1 p^{n-1} + \ldots + a_n}.$$

To do this we express the right-hand side of the equation in partial fractions.* Then

$$L[x(t)] = \frac{R(p)}{(p-p_1)(p-p_2)(p-p_3)\ldots(p-p_n)} = \tag{A.6}$$
$$= \frac{A_1}{p-p_1} + \frac{A_2}{p-p_2} + \ldots + \frac{A_n}{p-p_n},$$

where $p_1, p_2, \ldots, p_n$ ere the roots of the characteristic equation $D(p) = 0$; $A_1, A_2, \ldots, A_n$ are the coefficients (numerators) of the partial fractions.

We derive a formula for determining all the A. To do this we multiply both sides of the equation by $(p-p_k)$, where p_k is one of the roots of the equation $D(p) = 0$:

$$\frac{p-p_k}{p-p_1}A_1 + \frac{p-p_k}{p-p_2}A_2 + \ldots + \frac{p-p_k}{p-p_{k-1}}A_{k-1} + A_k +$$
$$+ \frac{p-p_k}{p-p_{k+1}}A_{k+1} + \ldots + \frac{p-p_k}{p-p_n}A_n = \frac{p-p_k}{D(p)}R(p).$$

In the resulting equation we put $p = p_k$. On the left only the term A_k is different from zero. On the right the fraction $\frac{p-p_k}{D(p)}$, when $p = p_k$, is of the indeterminate form $\frac{0}{0}$. Expanding it by l'Hôpital's rule we obtain $\lim\limits_{p\to p_k} \frac{p-p_k}{D(p)} = \frac{1}{D'(p_k)}$, so that the equation giving any coefficient A_k is of the form

$$A_k = \frac{R(p_k)}{D'(p_k)},$$

where

$$D'(p_k) = \left[\frac{\mathrm{d}D(p)}{\mathrm{d}p}\right]_{p=p_k}.$$

* We only consider the case when there are no multiple roots.

Putting these values of A_k in (A.6) we find an expression for $L[x(t)]$:

$$L[x(t)] = \frac{R(p)}{D(p)} = \frac{R(p_1)}{D'(p_1)} \frac{1}{p - p_1} + \\ + \frac{R(p_2)}{D'(p_2)} \frac{1}{p - p_2} + \ldots + \frac{R(p_n)}{D'(p_n)} \frac{1}{p - p_n},$$

or

$$L[x(t)] = \sum_{k=1}^{k=n} \frac{R(p_k)}{D'(p_k)(p - p_k)}. \qquad \text{(A.7)}$$

In each term the coefficient $\dfrac{R(p_k)}{D'(p_k)} = \text{const}$. In table A.1 above an expression for the original of $L[f(t)] = \dfrac{1}{p - a}$ was given. Using this expression, we find:

$$\frac{1}{p - p_k} \doteq e^{p_k t}.$$

Hence

$$\frac{R(p_k)}{D'(p_k)(p - p_k)} \doteq \frac{R(p_k)}{D'(p_k)} e^{p_k t}.$$

Thus, when $f(t) \equiv 0$ and when the initial conditions are different from zero the integral of equation (A.1) is equal to

$$x(t) = \sum_{k=1}^{k=n} \frac{R(p_k)}{D'(p_k)} e^{p_k t}. \qquad \text{(A.8)}$$

To find this integral the following steps are necessary.

(1) Determine $D(p)$, the left-hand side of the characteristic equation.

(2) Find the roots of this equation.

(3) Determine the coefficients of the polynomial $R(p)$ from (A.4).

(4) Find the derivative $D'(p)$;

(5) Put $p_1, p_2, \ldots, p_n$ successively in $D'(p)$ and in $R(p)$, and find the values of $\dfrac{R(p_k)}{D'(p_k)}$ (where $k = 1, 2, 3, 4, \ldots, n$);

(6) Form the sum (A.8).

If among the roots there are complex conjugates:

$$p_k = a + i\beta\,, \quad p_{k+1} = a - i\beta\,,$$

then

$$\frac{R(p_k)}{D'(p_k)} = \frac{\zeta + i\eta}{\lambda + i\xi} = \delta + i\sigma$$

and

$$\frac{R(p_{k+1})}{D'(p_{k+1})} = \frac{\zeta - i\eta}{\lambda - i\xi} = \delta - i\sigma\,.$$

We write complex numbers in vector form:

$$\frac{R(p_k)}{D'(p_k)} = Ae^{i\varphi}\,,$$

where

$$A = \sqrt{\delta^2 + \sigma^2}\,,\ \varphi = \arctan\frac{\sigma}{\delta}\,,$$

Then its complex conjugate is

$$\frac{R(p_{k+1})}{D'(p_{k+1})} = Ae^{-i\varphi}$$

and among the terms in

$$\sum_{k=1}^{k=n} \frac{R(p_k)}{D'(p_k)}\, e^{p_k t}$$

there will be terms

$$A\, e^{i\varphi}\, e^{(a+i\beta)t} + Ae^{-i\varphi}\, e^{(a-i\beta)t}\,.$$

Making use of Euler's identity

$$e^{iz} = \cos z + i \sin z\,,$$

we obtain

$$e^{i(\beta t+\varphi)} = \cos(\beta t + \varphi) + i\sin(\beta t + \varphi)\,,$$
$$e^{-i(\beta t+\varphi)} = \cos(\beta t + \varphi) - i\sin(\beta t + \varphi)\,.$$

Hence

$$Ae^{i\varphi}\, e^{(\alpha+i\beta)t} + Ae^{-i\varphi}\, e^{(\alpha-i\beta)t} =$$
$$= Ae^{\alpha t}\left[e^{i(\beta t+\varphi} + e^{-i(\beta t+\varphi)}\right] = 2\, Ae^{\alpha t} \cos\left(\beta t + \varphi\right).$$

Thus, when the characteristic equation has complex roots

$$x\left(t\right) = \sum_{k=1}^{k=r} C_k\, e^{p_k k} + \sum_{k=1}^{k=s} 2A_k\, e^{\alpha_k t} \cos\left(\beta_k\, t + \varphi_k\right), \tag{A.9}$$

where r is the number of real roots of the characteristic equation, and s is the number of pairs of complex conjugate roots of the characteristic equation,

$$C_k = \frac{R\left(p_k\right)}{D'\left(p_k\right)}, \quad A_k = \sqrt{\delta^2 + \sigma^2}\,,\ \varphi_k = \operatorname{arc\,tan} \frac{\sigma}{\delta}\,,$$

p_k is a real root of the characteristic equation, α_k and β_k are respectively the real and imaginary parts of the complex roots of the characteristic equation, and δ and σ are respectively real and imaginary parts of the expression $\dfrac{R\left(p_k\right)}{D'(p_k)}$ when p_k is a complex root.

This method of constructing the integral of a differential equation enables us to take the initial conditions into account from the very start, and simplifies the determination of the arbitrary constants, amplitudes and phases of the separate harmonics.

EXAMPLE 2. We are given the differential equation

$$\frac{\mathrm{d}^3\, x\left(t\right)}{\mathrm{d}t^3} + 6\, \frac{\mathrm{d}^2\, x\left(t\right)}{\mathrm{d}t^2} + 11\, \frac{\mathrm{d}x\left(t\right)}{\mathrm{d}t} + 6x\left(t\right) = 0$$

and, when $t = 0$, the initial conditions:

$$x\left(0\right) = 5,\ \frac{\mathrm{d}x\left(0\right)}{\mathrm{d}t} = \frac{\mathrm{d}^2\, x\left(0\right)}{\mathrm{d}t^2} = 0.$$

The characteristic equation

$$D(p) = p^3 + 6p^2 + 11p + 6 = 0$$

has the roots

$$p_1 = -1\,, \quad p_2 = -2\,, \quad p_3 = -3\,.$$

We calculate the coefficients of the polynomial

$$R(p) = b_1 p^2 + b_2 p + b_3, \quad b_1 = a_0 x(0) = 1 \times 5 = 5\,,$$

$$b_2 = a_0 \frac{dx(0)}{dt} + a_1 x(0) = 30,$$

$$b_3 = a_0 \frac{d^2 x(0)}{dt^2} + a_1 \frac{dx(0)}{dt} + a_2 x(0) = 55.$$

In this case

$$R(p) = 5p^2 + 30p + 55\,, \quad D'(p) = 3p^2 + 12p + 11$$

and

$$L[x(t)] = \frac{5p^2 + 30p + 55}{p^3 + 6p^2 + 11p + 6}\,.$$

Therefore

$$x(t) = \sum_{k=1}^{k=3} \frac{R(p_k)}{D'(p_k)} e^{p_k t} = \frac{5(1-6+11)}{3-12+11} e^{-t} +$$

$$+ \frac{5(4-12+11)}{12-24+11} e^{-2t} + \frac{5(9-18+11)}{27-36+11} e^{-2t} =$$

$$= 15e^{-t} - 15e^{-2t} + 5e^{-3t}.$$

Suppose now that $f(t) \not\equiv 0$, but that all the initial conditions are equal to zero, so that $R(p) \equiv 0$. Then

$$L[x(t)] = \frac{L[f(t)]}{D(p)}\,.$$

Usually in control theory the functions $f(t)$ are those given in table A.1. All these functions have transforms of fractional-rational form $\frac{r(p)}{s(p)}$ where $r(p)$ and $s(p)$ are polynomials in p of the second, first or zero degree (see table A.1).

Thus

$$L[x(t)] = \frac{r(p)}{s(p)\,D(p)} = \frac{r(p)}{\overline{D}(p)},$$

where

$$\overline{D}(p) = s(p)\,D(p)\,.$$

Reasoning in exactly the same way as we did above in the case $f(t) \equiv 0$, $R(p) \neq 0$ we obtain the required original $x(t)$ in this case:

$$x(t) = \sum_k \frac{r(p_k)}{\overline{D}'(p_k)}\,e^{p_k t}, \tag{A.10}$$

where p_k are the roots of the equation $\overline{D}(p) = 0$. The summation is taken over all the roots.

If $D(p)$ has one zero root, i.e. $\overline{D}(p) = p\overline{D}_1(p)$, formula (A.10) can be put in the form

$$x(t) = \frac{r(0)}{\overline{D}_1(0)} + \sum_k \frac{r(p_k)}{p_k\,\overline{D}_1'(p_k)}\,e^{p_k t}$$

Example 3. We consider the differential equation

$$\frac{d^3 x(t)}{dt^3} + 6\,\frac{d^2 x(t)}{dt^2} + 11\,\frac{dx(t)}{dt} + 6x(t) = \mathbf{1}\,(1 + e^{-4t}).$$

The initial conditions are: $x(0) = \dfrac{dx(0)}{dt} = \dfrac{d^2 x(0)}{dt^2} = 0$. Due to the fact that all the initial conditions are equal to zero, $R(p)$ is identically equal to zero.

From table A.1

$$L[f(t)] = \left(\frac{1}{p} + \frac{1}{p+4}\right) = \frac{2p+4}{p(p+4)}\,.$$

In this case

$$\begin{aligned}
r(p) &= 2p + 4,\\
s(p) &= p(p+4),\\
D(p) &= p^3 + 6p^2 + 11p + 6,\\
\overline{D}(p) &= p^5 + 10p^4 + 35p^3 + 50p^2 + 24p,\\
\overline{D}'(p) &= 5p^4 + 40p^3 + 105p^2 + 100p + 24.
\end{aligned}$$

Hence

$$L\,[x\,(t)] = \frac{2p+4}{p\,(p+4)\,(p^3+6p^2+11p+6).}\,.$$

The roots of the characteristic equation $D(p) = 0$ are equal to -1, -2 and -3.

The roots of the equation $s(p) = 0$ are equal to 0 and -4.

The original $x(t)$ is equal to

$$x\,(t) = \frac{4}{24} - \frac{1}{6}\,e^{-4t} + \frac{1}{3}\,e^{-3t} - \frac{1}{3}\,e^{-t}.$$

Let us now assume that $f(t) \neq 0$ and that not all the initial conditions are zero. Then the transform $L\,[x(t)]$ contains the two terms:

$$\frac{L\,[f\,(t)]}{D\,(p)} \text{ and } \frac{R\,(p)}{D\,(p)}\,,$$

and the original function is equal to the sum of the inverse transforms of these functions.

EXAMPLE 4. We consider the differential equation

$$\frac{\mathrm{d}^3 x}{\mathrm{d}t^2} + 6\,\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + 11\,\frac{\mathrm{d}x}{\mathrm{d}t} + 6x = 1\,(1 - e^{-4t})$$

with the initial conditions, when $t = 0$: $x = 5$, $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{\mathrm{d}^2 x}{\mathrm{d}t^2} = 0$.

In example 2 the original of $\dfrac{R\,(p)}{D(p)}$ was found for this case, and the original of $\dfrac{L\,[f\,(t)]}{D\,(p)}$ was found in example 3. To find the integral of the given equation, we must add the solutions of examples 2 and 3.

3. The Integration of a System of Linear Differential Equations

Usually the processes of automatic control are described not by one but by a system of equations, in the general case by a system of the form

$$\left[a_{11}\,\frac{\mathrm{d}^2 x_1\,(t)}{\mathrm{d}t^2} + b_{11}\,\frac{\mathrm{d}x_1\,(t)}{\mathrm{d}t} + c_{11}\,x_1\,(t)\right] +$$

$$+\left[a_{12}\frac{d^2 x_2(t)}{dt^2}+b_{12}\frac{dx_2(t)}{dt}+c_{12}x_2(t)\right]+\ldots$$

$$\ldots+\left[a_{1n}\frac{d^2 x_n(t)}{dt^2}+b_{1n}\frac{dx_n(t)}{dt}+c_{1n}x_n(t)\right]=f_1(t),$$

$$\left[a_{21}\frac{d^2 x_1(t)}{dt^2}+b_{21}\frac{dx_1(t)}{dt}+c_{21}x_1(t)\right]+$$

$$+\left[a_{22}\frac{d^2 x_2(t)}{dt^2}+b_{22}\frac{dx_2(t)}{dt}+c_{22}x_2(t)\right]+\ldots$$

$$\ldots+\left[a_{2n}\frac{d^2 x_n(t)}{dt^2}+b_{2n}\frac{dx_n(t)}{dt}+c_{2n}x_n(t)\right]=f_2(t),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\left[a_{n1}\frac{d^2 x_1(t)}{dt^2}+b_{n1}\frac{dx_1(t)}{dt}+c_{n1}x_1(t)\right]+$$

$$+\left[a_{n2}\frac{d^2 x_2(t)}{dt^2}+b_{n2}\frac{dx_3(t)}{dt}=c_{n2}x_2(t)\right]+\ldots$$

$$\ldots+\left[a_{nn}\frac{d^2 x_n(t)}{dt^2}+b_{nn}\frac{dx_n(t)}{dt}+c_{nn}x_n(t)\right]=f_n(t),$$

$$\text{(A.11)}$$

where $x_1(t)$, $x_2(t)$, ... , $x_n(t)$ are the required functions of t; a, b, c, are real numbers (in every real case some of these numbers will be zero) and $f_i(t)$ are functions having values different from zero when $t \geqslant 0$.

To determine the law of change of any of the generalized coordinates $x_k(t)$ we could, by eliminating the other coordinates, replace this system of differential equations by a single equation in x_k, and, further, replace the initial conditions given for the system (A.11) by corresponding conditions for this equation and then work according to the rules given above. It is simpler, however, when a system of linear differential equations is given, to find the transform of each equation separately in the same way as was done above for one equation. To do this we must multiply the left- and right-hand sides of each equation by e^{-pt} and then integrate them between 0 and ∞.

As a result we obtain a system of algebraic equations in terms of the transforms $L[x_1(t)]$, $L[x_2(t)]$, $\ldots$, $L[x_n(t)]$:

$$\left.\begin{array}{l} D_{11}(p)\,L[x_1(t)] + D_{12}(p)\,L[x_2(t)] + \ldots \\ \ldots + D_{1n}(p)\,L[x_n(t)] = L[f_1(t)] + R_1(p), \\ D_{21}(p)\,L[x_1(t)] + D_{22}(p)\,L[x_2(t)] + \ldots \\ \ldots + D_{2n}(p)\,L[x_n(t)] = L[f_2(t)] + R_2(p), \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ D_{n1}(p)\,L[x(t)] + D_{n2}(p)\,L[x_2(t)] + \ldots \\ \ldots + D_{nn}(p)\,L[x_n(t)] = L[f_n(t)] + R_n(p), \end{array}\right\} \qquad (A.12)$$

where all the $R_1(p)$, $R_2(p)$, $\ldots$, $R_n(p)$ are polynomials in p with coefficients which depend on the initial conditions,* and all the $D_{ij}(p)$ are determined as in the case of a single differential equation.

The system (A.12) contains n linear algebraic equations in n unknowns:

$$L[x_1(t)]\,,\quad L[x_2(t)]\,,\;\ldots\;, L[x_n(t)]\,.$$

If we are interested in any one coordinate, x_1 for example, then its transform can be found by solving the system (A.12) with respect to $L[x_1] = \dfrac{\Delta_1}{\Delta}$ where:

$$\Delta = \begin{vmatrix} D_{11}(p) & D_{12}(p) \ldots D_{1n}(p) \\ D_{21}(p) & D_{22}(p) \ldots D_{2n}(p) \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ D_{n1}(p) & D_{n2}(p) \ldots D_{nn}(p) \end{vmatrix},$$

$$\Delta_1 = \begin{vmatrix} L[f_1(t)] + R_1(p) & D_{12}(p) \ldots D_{1n}(p) \\ L[f_2(t)] + R_2(p) & D_{22}(p) \ldots D_{2n}(p) \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ L[f_n(t)] + R_n(p) \ldots & D_{n2}(p) \ldots D_{nn}(p) \end{vmatrix}.$$

* The functions $R_1(p)$, $R_2(p)$, $\ldots$, $R_n(p)$ are determined during the transition from the system (A.11) to (A.12). They cannot be calculated directly from formulae (A.3) and (A.4) which were given above for the calculation of $R(p)$ in the case of the transform of a single differential equation, since in this case each of the equations in the system (A.11) contains several of the required functions $x_j(t)$.

Expanding the determinant Δ_1, we obtain:

$$L[x_1(t)] = \frac{\Delta_1}{\Delta} = \qquad (A.13)$$

$$= \frac{L[f_1(t)]\Delta_{11} + \ldots + L[f_n(t)]\Delta_{1n} + R_1(p)\Delta_{11} + \ldots + R_n(p)\Delta_{1n}}{D(p)},$$

where $D(p) = \Delta(p)$ is the left-hand side of the characteristic equation of the system, and $\Delta_{11}, \Delta_{12}, \ldots, \Delta_{1n}$ are algebraic complements of the corresponding elements in the first column.

If

$$f_1(t) = f_2(t) = \ldots = f_n(t) = 0,$$

then, also

$$L[f_1(t)] = L[f_2(t)] = \ldots = L[f_n(t)] = 0.$$

In this case $L(x_1)$ once again reduces to the form

$$L[x_1] = \frac{\overline{R}(p)}{D(p)},$$

where $\overline{R}(p) = R_1(p)\Delta_{11} + \ldots + R_n(p)\Delta_{1n}$ is a polynomial in p whose coefficients depend on the initial conditions. Hence, in this case also

$$x_1(t) = \sum_{k=1}^{k=n} \frac{\overline{R}(p_k)}{D'(p_k)} e^{p_k t}. \qquad (A.14)$$

If all the initial conditions are zero, but

$$f_i(t) \neq 0,$$

then

$$L[x_1(t)] = \frac{M(p)}{D(p)}, \qquad (A.15)$$

where

$$M(p) = L[f_1(t)]\Delta_{11} + \ldots + L[f_n(t)]\Delta_{1n}.$$

When all the $L[f_j(t)]$ are fractional-rational functions (see table A.1) $M(p)$ can be reduced to the form $M(p) = \dfrac{r(p)}{s(p)}$ and we obtain once again

$$L[x_1] = \frac{r(p)}{s(p)D(p)}. \qquad (A.16)$$

Then $x_1(t)$ is given by formula (A.10).

EXAMPLE 5. To find the function $x_1(t)$ which satisfies the differential equations

$$\left.\begin{aligned} \frac{dx_1(t)}{dt} + x_1(t) + 5x_2(t) &= \mathit{1}, \\ 0.1\frac{dx_2(t)}{dt} + x_2(t) - 10x_1(t) &= 0. \end{aligned}\right\}$$

with the initial conditions $x_1(0) = x_2(0) = \frac{dx_1(0)}{dt} = \frac{dx_2(0)}{dt} = 0.$
We take the Laplace transform:

$$\begin{aligned} (p+1)L[x_1(t)] + 5L[x_2(t)] &= L(\mathit{1}), \\ -10L[x_1(t)] + (0.1p+1)L[x_2(t)] &= 0. \end{aligned}$$

In this case

$$L[x_1(t)] = \frac{\begin{vmatrix} L[\mathit{1}] & 5 \\ 0 & 0.1p+1 \end{vmatrix}}{\begin{vmatrix} p+1 & 5 \\ -10 & 0.1p+1 \end{vmatrix}} = \frac{(0.1p+1)L[1]}{(p+1)(0.1p+1)+50}.$$

From table A.1 it is seen that the transform of $L[\mathit{1}] = \frac{1}{p}$ and

$$L[x_1(t)] = \frac{0.1p+1}{p(0.1p^2+1.1p+51)}.$$

In this case

$$\begin{aligned} r(p) &= 0.1p+1, \\ s(p) &= p, \\ D(p) &= 0.1p^2+1.1p+51, \\ \overline{D(p)} &= 0.1p^3+1.1p^2+51p, \\ \overline{D'(p)} &= 0.3p^2+2.2p+51. \end{aligned}$$

The roots of $D(p) = 0$ are equal to $p_{1.2} = -5.5 \pm i\,21.9$.
The roots of $s(p) = 0$ is equal to $p_3 = 0$.

We obtain the required original by using formula (A.10):

$$x_1(t) = \frac{1}{51} + 0{\cdot}0466\, e^{-5{\cdot}5t} \cos(21{\cdot}9t + 115°43') =$$
$$= 0{\cdot}0106 - 0{\cdot}0466\, e^{-5{\cdot}5t} \sin(21{\cdot}9t + 25°43').$$

Similarly we can find $x_2(t)$:

$$L[x_2(t)] = \frac{\begin{vmatrix} p+1 & L[1] \\ -10 & 0 \end{vmatrix}}{\begin{vmatrix} p+1 & 5 \\ -10 & 0{\cdot}1p+1 \end{vmatrix}} =$$
$$= \frac{+10L[1]}{(p+1)(0{\cdot}1p+1)+50} = \frac{10}{p[(p+1)(0{\cdot}1p+1)+50]}.$$

The roots of $D(p) = 0$ are equal to $p_{1.2} = 5{\cdot}5 \pm i\, 21{\cdot}9$, and the root of $s(p) = 0$ is equal to $p_3 = 0$.

From formula (A.10) we find:

$$x_2(t) = \frac{10}{51} + 0{\cdot}202\, e^{-5{\cdot}5t} \cos(21{\cdot}9t + 194°6') =$$
$$= 0{\cdot}196 - 0{\cdot}202\, e^{-5{\cdot}5t} \cos(21{\cdot}9t + 14°6').$$

4. The Fourier Transform and Integral

The Fourier transform is a special case of the Laplace transform. It is obtained from the Laplace transform by putting $p = i\omega$. Thus, we obtain an expression defining the Fourier transform in the form

$$\Phi[f(t)] = \int_0^\infty f(t)\, e^{-i\omega t}\, \mathrm{d}t. \tag{A.17}$$

Stricter restrictions are laid on the function when the Fourier transform is used than in the use of the Laplace transform.* If these

* For more detail see the literature given in the bibliography to the Appendix.

restrictions have been taken into account all the properties and tables formed for the Laplace transform can also be extended to this transform. It is only necessary to put $p = i\omega$ in them.

In contrast to the Laplace transform the Fourier transform enables us to give an intuitive physical interpretation of the process of determining the transform and its original. To explain this property of the Fourier Transform, let us examine some properties of Fourier series and of the Fourier integral.

Let any periodic function $f(t)$ with period T which satisfies the usual conditions necessary for an expansion in a Fourier series be given:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos 2\pi \frac{k}{T} t + b_k \sin 2\pi \frac{k}{T} t \right). \tag{A.18}$$

where $\dfrac{2\pi}{T} = \omega$ is the frequency. The coefficients of this expansion are defined by the formulae:

$$a_0 = \frac{2}{T} \int_0^T f(t)\, dt,$$

$$a_k = \frac{2}{T} \int_0^T f(t) \cos \frac{2\pi}{T} kt\, dt,$$

$$b_k = \frac{2}{T} \int_0^T f(t) \sin \frac{2\pi}{T} kt\, dt.$$

The expression (A.18) can be written:

$$f(t) = \frac{a_0}{2} + \sum A_k \cos (k\omega t - \varphi_k), \tag{A.19}$$

where

$$A_k = \sqrt{(a_k^2 + b_k^2)},$$

$$\varphi_k = \text{arc tan} \frac{b_k}{a_k},$$

$$\omega = \frac{2\pi}{T}.$$

We consider the general term of the Fourier series:

$$a_k \cos \frac{2\pi}{T} kt + c_k \sin \frac{2\pi}{T} kt.$$

Using the formulae which connect the trigonometric functions and the hyperbolic functions:

$$\cos \tau = \cosh i\tau = \frac{1}{2} (e^{i\tau} + e^{-i\tau}),$$

$$\sin \tau = \frac{1}{i} \sinh i\tau = \frac{1}{2i} (e^{i\tau} - e^{-i\tau}),$$

where

$$\tau = \frac{2\pi}{T} kt.$$

Then, putting these values of $\cos \frac{2\pi}{T} kt$ and $\sin \frac{2\pi}{T} kt$ in the general term of the Fourier series, we obtain:

$$\frac{a_k}{2} (e^{i\tau} + e^{-i\tau}) + \frac{b_k}{2i} (e^{i\tau} - e^{-i\tau}).$$

Collecting all the terms containing $e^{i\tau}$ and $e^{-i\tau}$, we find:

$$\frac{1}{2} (a_k - ib_k) e^{i\tau} + \frac{1}{2} (a_k + ib_k) e^{-i\tau}.$$

Let:

$$\frac{1}{2} (a_k - ib_k) = C_k, \quad \frac{1}{2} (a_k + ib_k) = C_{-k}, \quad \frac{a_0}{2} = C_0.$$

Then

$$a_k \cos \frac{2\pi}{T} kt + b_k \sin \frac{2\pi}{T} kt = C_k e^{i\frac{2\pi}{T}kt} + C_{-k} e^{-i\frac{2\pi}{T}kt}.$$

Thus, the Fourier expansion of the periodic function $f(t)$ can be written in the complex form:

$$f(t) = \sum_{k=-\infty}^{k=+\infty} C_k e^{i\frac{2\pi}{T}kt}. \tag{A.20}$$

The coefficients of the series are determined by the formula

$$C_k = \frac{1}{T}\int_0^T f(t)\, e^{-i\frac{2\pi}{T}kt}\, dt.$$

Indeed,

$$C_k = \frac{a_k - ib_k}{2} = \frac{1}{T}\int_0^T f(t)\left[\cos\frac{2\pi}{T}kt - i\sin\frac{2\pi}{T}kt\right]dt =$$

$$= \frac{1}{T}\int_0^T f(t)\, e^{-i\frac{2\pi}{T}kt}\, dt.$$

If we know C_k it is easy to compute the amplitude and phase of the oscillation for any k. k can only take integral values.

The graph on which, for each integral value of k the value of C_k is drawn, is called the spectrum.

Thus, a Fourier series enables us to represent a periodic function in the form of a discrete spectrum. If the function $f(t)$ is unperiodic and is given in some finite interval of time (and has no meaning when it is outside this interval) we can take this interval as the period T and expand the function in a Fourier series.

As $T \to \infty$ the series becomes the Fourier integral

$$f(t) = \int_0^\infty [a(\omega)\cos\omega t + b(\omega)\sin\omega t]\, d\omega.$$

The coefficients are determined by means of passing to the limit as $T \to \infty$ in the general expression for a_k and b_k:

$$a_k = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\cos\omega t\, dt = a(\omega), \quad b_k = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\sin\omega t\, dt = b(\omega).$$

The Fourier integral can be written:

$$f(t) = \int_0^\infty A(\omega)\sin[\omega t + \varphi(\omega)]\, d\omega,$$

where

$$A(\omega) = \sqrt{a^2(\omega) + b^2(\omega)}, \quad \varphi(\omega) = \arctan\frac{a(\omega)}{b(\omega)}.$$

Putting $A(\omega)$ and $\varphi(\omega)$ along the ordinate axis, and taking ω as the abscissa, we obtain continuous curves, the continuous spectrum of an unperiodic function.

Let us write the Fourier integral in a more compact complex form. We note that

$$b(\omega)\sin\omega t + a(\omega)\cos\omega t =$$
$$= \frac{b(\omega)}{2i}(e^{i\omega t} - e^{-i\omega t}) + \frac{a(\omega)}{2}(e^{i\omega t} + e^{-i\omega t}) =$$
$$= \frac{a(\omega) - ib(\omega)}{2} e^{i\omega t} + \frac{a(\omega) + ib(\omega)}{2} e^{-i\omega t}.$$

Putting $F_1(i\omega) = \dfrac{a(\omega) - ib(\omega)}{2}$, we have

$$f(t) = \int_{+\infty}^{=\infty} F_1(i\omega)\, e^{i\omega t}\, d\omega,$$

where

$$F_1(i\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt.$$

If we put $\int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt = F(i\omega)$, then

$$F_1(i\omega) = \frac{1}{2\pi} F(i\omega).$$

Thus, the unperiodic function $f(t)$ (satisfying restrictions which we have not mentioned) can be represented in the form of the Fourier integral:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(i\omega)\, e^{i\omega t}\, d\omega,$$

where

$$F(i\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt.$$

The integral $F(i\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt$ itself represents the Fourier transform of the function $f(t)$. The lower limit can be replaced by zero if in the Fourier transform, just as in the Laplace transform, only those functions which are equal to zero for $t < 0$ are considered. We can call the function $F(i\omega)$ the complex spectrum of the function $f(t)$.

APPENDIX 2

Table of the trigonometric functions for angles expressed in radians, and of the functions e^{-x}, $\frac{\sin x}{x}$, $\frac{\cos x}{x}$ and Si x, the integral sine

x	Si x	$\cos x$	$\sin x$	e^{-x}	$\frac{\sin x}{x}$	$\frac{\cos x}{x}$
0·00	0	1	0	1	1	∞
0·01	0·010000	0·99995	0·01000	0·99005	0·99998	00·99500
0·02	0·019999	0·99980	0·02000	0·98020	0·99993	49·9900
0·03	0·029998	0·99955	0·03000	0·97045	0·99985	33·31833
0·04	0·039996	0·99920	0·03999	0·96079	0·99973	24·98000
0·05	0·04999	0·99875	0·04998	0·95123	0·99958	19·97500
0·06	0·05999	0·99820	0·05996	0·94176	0·99940	16·63667
0·07	0·06998	0·99775	0·06994	0·93239	0·99918	14·25072
0·08	0·07997	0·99680	0·07991	0·92312	0·99893	12·46002
0·09	0·08996	0·99595	0·08988	0·91393	0·99865	11·06614
0·10	0·09994	0·99500	0·09983	0·90484	0·99833	9·95004
0·11	0·10993	0·99396	·010978	0·89583	0·99798	9·05596
0·12	0·11990	0·99281	0·11971	0·88692	0·99760	8·27340
0·13	0·12998	0·99156	0·12963	0·87810	0·99718	7·62739
0·14	0·13985	0·99022	0·13954	0·86936	0·99673	7·07297
0·15	0·14981	0·98877	0·14944	0·86071	0·99625	6·59180
0·16	0·15977	0·98723	0·15932	0·85214	0·99573	6·17017
0·17	0·16973	0·98558	0·16918	0·84366	0·99519	5·79755
0·18	0·1797	0·98384	0·17903	0·83527	0·99460	5·46579
0·19	0·1896	0·98200	0·18886	0·82696	0·99399	5·16844
0·20	0·1996	0·98007	0·19867	0·81873	0·99334	4·90033
0·21	0·2095	0·97803	0·20846	0·81058	0·99266	4·65729
0·22	0·2194	0·97590	0·21823	0·80252	0·99195	4·43589
0·23	0·2293	0·97367	0·22798	0·79453	0·99120	4·23333
0·24	0·2392	0·97134	0·23770	0·78663	0·99042	4·04724
0·25	0·2491	0·96891	0·24740	0·77880	0·98961	3·87561
0·26	0·2590	0·96639	0·25708	0·77105	0·98877	3·71688
0·27	0·2689	0·96377	0·26673	0·76338	0·98789	3·56952
0·28	0·2788	0·96106	0·27636	0·75578	0·98698	3·43234
0·29	0·2886	0·95824	0·28595	0·74826	0·98604	3·30428
0·30	0·2985	0·95534	0·29552	0·74082	0·98506	3·18445
0·31	0·3083	0·95233	0·30506	0·73345	0·98406	3·07204
0·32	0·3182	0·94924	0·31457	0·72615	0·98302	2·96636
0·33	0·3280	0·94604	0·32404	0·71892	0·98194	2·86679
0·34	0·3378	0·94275	0·33349	0·71177	0·98084	2·77280
0·35	0·3476	0·93937	0·34290	0·70469	0·97970	2·68392

x	Si x	$\cos x$	$\sin x$	e^{-x}	$\frac{\sin x}{x}$	$\frac{\cos x}{x}$
0·36	0·3574	0·93590	0·35227	0·69768	0·97853	2·59971
0·37	0·3672	0·93233	0·36162	0·69073	0·97733	2·51980
0·38	0·3770	0·92866	0·37092	0·68386	0·97610	2·44385
0·39	0·3867	0·92491	0·38019	0·67706	0·97484	2·37156
0·40	0·3965	0·92016	0·38942	0·67032	0·97354	2·30265
0·41	0·4062	0·91712	0·39861	0·66365	0·97221	2·23688
0·42	0·4159	0·91309	0·40776	0·65705	0·97085	2·17402
0·43	0·4256	0·90897	0·41687	0·65051	0·96946	2·11387
0·44	0·4353	0·90475	0·42594	0·64404	0·96804	2·05625
0·45	0·4450	0·90045	0·43497	0·63763	0·96659	2·00099
0·46	0·4546	0·89605	0·44395	0·63128	0·96510	1·94794
0·47	0·4643	0·89157	0·45289	0·62500	0·96358	1·89695
0·48	0·4739	0·88699	0·46178	0·61878	0·96203	1·84790
0·49	0·4835	0·88233	0·47063	0·61263	0·96046	1·80067
0·50	0·4931	0·87758	0·47943	0·60653	0·95855	1·75516
0·51	0·5027	0·87274	0·48818	0·60050	0·95721	1·71126
0·52	0·5123	0·86782	0·49688	0·59452	0·95553	1·66888
0·53	0·5218	0·86281	0·50559	0·58860	0·95383	1·62793
0·54	0·5313	0·85771	0·51414	0·58275	0·95210	1·58834
0·55	0·5408	0·85252	0·52268	0·57695	0·95034	1·55004
0·56	0·5503	0·84726	0·53119	0·57121	0·94854	1·51295
0·57	0·5598	0·84190	0·53963	0·56553	0·94672	1·47701
0·58	0·5693	0·83646	0·54802	0·55990	0·94486	1·44217
0·59	0·5787	0·83094	0·55636	0·55433	0·94298	1·40837
0·60	0·5881	0·82534	0·56464	0·54881	0·94107	1·37555
0·61	0·5975	0·81965	0·57287	0·54335	0·93912	1·34368
0·62	0·6069	0·81388	0·58104	0·53794	0·93715	1·31270
0·63	0·6163	0·80803	0·58914	0·53259	0·93515	1·28258
0·64	0·6256	0·80210	0·59720	0·52729	0·93311	1·25327
0·65	0·6349	0·79608	0·60519	0·52205	0·93105	1·22474
0·66	0·6442	0·78999	0·61312	0·51685	0·92876	1·19695
0·67	0·6435	0·78382	0·62099	0·51171	0·92684	1·16988
0·68	0·6628	0·77757	0·62879	0·50662	0·92469	1·14348
0·69	0·6720	0·77125	0·63654	0·50158	0·92251	1·11774
0·70	0·6812	0·76484	0·64422	0·49659	0·92031	1·09263
0·71	0·6904	0·75836	0·65183	0·49164	0·91807	1·06811
0·72	0·6996	0·75181	0·65938	0·48675	0·91581	1·04417
0·73	0·7087	0·74517	0·66687	0·48191	0·91352	1·02078
0·74	0·7179	0·73847	0·67429	0·47711	0·91119	0·99793
0·75	0·7270	0·73169	0·68164	0·47237	0·90885	0·97558
0·76	0·7360	0·72484	0·68892	0·46767	0·90647	0·95373
0·77	0·7451	0·71791	0·69614	0·46301	0·90407	0·93235

x	Si x	cos x	sin x	e^{-x}	$\frac{\sin x}{x}$	$\frac{\cos x}{x}$
0·78	0·7541	0·71091	0·70328	0·45841	0·90164	0·91142
0·79	0·7631	0·70385	0·71035	0·45384	0·89918	0·89094
0·80	0·7721	0·69671	0·71736	0·44933	0·89669	0·87088
0·81	0·7811	0·68950	0·72429	0·44486	0·89418	0·85123
0·82	0·7900	0·68222	0·73115	0·44043	0·89164	0·83197
0·83	0·7989	0·67488	0·73793	0·43605	0·88907	0·81310
0·84	0·8078	0·66746	0·74464	0·43171	0·88647	0·79459
0·85	0·8166	0·65998	0·75128	0·42741	0·88385	0·77645
0·86	0·8254	0·65244	0·75784	0·42316	0·88121	0·75864
0·87	0·8342	0·64483	0·76433	0·41895	0·87853	0·74117
0·88	0·8430	0·63715	0·77074	0·41478	0·87583	0·72403
0·89	0·8518	0·62941	0·77707	0·41066	0·87311	0·70720
0·90	0·8605	0·62162	0·78333	0·40657	0·87036	0·69067
0·91	0·8692	0·61375	0·78950	0·40252	0·86758	0·67444
0·92	0·8778	0·60582	0·79560	0·39852	0·86478	0·65850
0·93	0·8865	0·59783	0·80162	0·39455	0·86195	0·64283
0·94	0·8951	0·58979	0·80756	0·39063	0·85910	0·62743
0·95	0·9036	0·58108	0·81342	0·38674	0·85622	0·61229
0·96	0·9122	0·57352	0·81919	0·38289	0·85332	0·59741
0·97	0·9207	0·56530	0·82489	0·37908	0·85039	0·58278
0·98	0·9292	0·55702	0·83050	0·37531	0·84744	0·56839
0·99	0·9377	0·54869	0·83603	0·37158	0·84447	0·55423
1·00	0·9461	0·54030	0·84147	0·36788	0·84147	0·54030
1·1	1·0287	0·45360	0·89121	0·33287	0·81018	0·41236
1·2	1·1080	0·36236	0·93204	0·30119	0·77669	0·30196
1·3	1·1840	0·26750	0·96356	0·27253	0·74119	0·20576
1·4	1·2562	0·16997	0·98545	0·24660	0·70389	0·12140
1·5	1·3247	0·07074	0·99749	0·22313	0·66499	0·04715
$\frac{\pi}{2}$	1·3699	0	1	0·20788	0·63662	0
1·6	1·3892	—0·02920	0·99957	0·20190	0·62473	—0·01824
1·7	1·4496	—0·12884	0·99166	0·18268	0·58333	—0·07579
1·8	1·5058	—0·22720	0·97385	0·16530	0·54102	—0·12622
1·9	1·5578	—0·32329	0·94630	0·14957	0·49805	—0·17015
2·0	1·6054	—0·41615	0·90930	0·13534	0·45464	—0·20807
2·1	1·6487	—0·50485	0·86321	0·12246	0·41105	—0·24040
2·2	1·6876	—0·58850	0·80850	0·11080	0·36749	—0·26750
2·3	1·7222	—0·66628	0·74571	0·10026	0·32421	—0·28968
2·4	1·7525	—0·73739	0·67546	0·09072	0·28144	—0·30724
2·5	1·7785	—0·80114	0·59847	0·08208	0·23938	—0·32046
2·6	1·8004	—0·85689	0·51550	0·07427	0·19827	—0·32957
2·7	1·8182	—0·90407	0·42738	0·06721	0·15829	—0·33484

x	Si x	$\cos x$	$\sin x$	e^{-x}	$\dfrac{\sin x}{x}$	$\dfrac{\sin x}{x}$
2·8	1·8321	—0·94222	0·33499	0·06081	0·11964	—0·33651
2·9	1·8422	—0·97096	0·23925	0·05502	0·08250	—0·33481
3·0	1·8487	—0·98999	0·14112	0·04979	0·04704	—0·32999
3·1	1·8517	—0·99914	0·04158	0·04505	0·01341	—0·32230
π	1·8519	—1	0	0·04321	0	—0·31831
3·2	1·8514	—0·99829	—0·05837	0·04076	—0·01824	—0·31196
3·3	1·8481	—0·98748	—0·15775	0·03688	—0·04780	—0·29923
3·4	1·8419	—0·96680	—0·25554	0·03337	—0·07515	—0·28435
3·5	1·8331	—0·93646	—0·35078	0·03020	—0·10022	—0·26755
3·6	1·8219	—0·89676	—0·44252	0·02732	—0·12292	—0·24909
3·7	1·8086	—0·84810	—0·52984	0·02472	—0·14319	—0·22921
3·7	1·7934	—0·79097	—0·61186	0·02237	—0·16101	—0·20814
3·9	1·7765	—0·72593	—0·68777	0·02024	—0·17635	—0·18613
4·0	1·7582	—0·65364	—0·75680	0·01832	—0·18920	—0·16341
4·1	1·7387	—0·57482	—0·81828	0·01657	—0·19958	—0·14020
4·2	1·7184	—0·49026	—0·87158	0·01500	—0·20752	—0·11673
4·3	1·6973	—0·40080	—0·91617	0·01357	—0·21306	—0·09321
4·4	1·6758	—0·30733	—0·95160	0·01228	—0·21627	—0·06985
4·5	1·6541	—0·21080	—0·97753	0·01111	—0·21722	—0·04684
4·6	1·6325	—0·11215	—0·99369	0·010052	—0·21601	—0·02438
4·7	1·6110	—0·01239	—0·99992	0·009095	—0·21274	—0·00263
1·5π	1·6089	0	—1	0·008983	—0·21221	0
4·8	1·5900	—0·08750	—0·99616	0·008230	—0·20753	0·01822
4·9	1·5696	—0·18651	—0·98245	0·007447	—0·20050	0·03806
5·0	1·5499	—0·28366	—0·95892	0·006738	—0·19178	0·05673
6	1·4247	—0·96017	—0·27942	0·002479	—0·04656	0·16002
2π	1·4182	1	0	0·001867	0	0 15915
7	1·4546	0·75390	0·65699	0·0009119	0·09386	0·10770
8	1·5742	—0·14550	0·98936	0·0003355	0·12366	—0·01818
9	1·6650	—0·91113	0·41212	0·0001234	0·04579	—0·10123
10	1·6583	—0·83907	—0·54402	0·0000460	—0·05440	—0·08390
11	1·5783	+0·00443	—0·9999	0·0000167	—0·09090	+0·00040
12	1·5050	+0·84385	—0·53657	0·0000061	—0·04471	0·07032
13	1·4994	+0·90745	+0·42017	0·0000023	0·03232	0·06980
14	1·5562	+0·13674	+0·99061		0·07075	0·00976
15	1·6182	—0·75969	+0·65029		0·04335	—0·05064
20	1·5482	+0·40808	+0·91295		0·04565	0·02040
25	1·5315	+0·99120	—0·13235		—0·00529	0·03965
30	1·5668	+0·15425	—0·98803		—0·03293	+0·00154
35	1·5969	—0·90369	—0·42818		—0·01223	—0·02582
40	1·5870	—0·66694	+0·74511		+0·01863	—0·01667
45	1·5587	+0·52532	+0·85090		0·01891	0·01167

x	Si x	$\cos x$	$\sin x$	e^{-x}	$\dfrac{\sin x}{x}$	$\dfrac{\cos x}{x}$
50	1·5516	+0·96497	—0·26237		—0·00525	0·01930
55	1·5707	+0·02210	—0·99976		—0·01818	0·00040
60	1·5867	—0·95243	—0·30487		—0·00508	—0·01587
65	1·5792	—0·56249	+0·82680		0·01272	—0·00865
70	1·5616	+0·63329	+0·77391		0·01106	+0·00905
75	1·5586	+0·92175	—0·38685		—0·00517	0·01229
80	1·5723	—0·11038	—0·99389		—0·01242	—0·00138
85	1·5824	—0·98438	—0·17608		—0·00207	—0·01158
90	1·5757	—0·44806	+0·89400		+0·00993	—0·00498
95	1·5630	+0·73017	+0·68326		0·00719	0·00769
100	1·5622	+0·86230	—0·50640		—0·00506	0·00862
110	1·5799	—0·99902	—0·04429		—0·00040	—0·00909
120	1·5640	+0·81417	—0·58063		—0·00484	+0·00679
130	1·5737	—0·36729	—0·93011		—0·00283	—0·00716

Illustrations: 1. The function Si x, as $x \to \infty$, tends to the quantity $\dfrac{\pi}{2}$.

2. The extrema of Si x are observed at $x = \pi, 2\pi, 3\pi, \ldots$ The difference between $x_{\max}$ (or $x_{\min}$) and $\dfrac{\pi}{2}$ is given in the following table:

$\dfrac{x}{\pi}$	$x_{\max} - \dfrac{\pi}{2}$ or $x_{\min} - \dfrac{\pi}{2}$	$\dfrac{x}{\pi}$	$x_{\max} - \dfrac{\pi}{2}$ or $x_{\min} - \dfrac{\pi}{2}$
1	+0·28114	9	+0·035280
2	—0·15264	10	—0·031767
3	+0·10396	11	+0·028889
4	—0·078635	12	—0·026489
5	+0·063168	13	+0·024456
6	—0·052762	14	—0·022713
7	+0·045289	15	+0·021201
8	—0·039665		

The function Si x is given by the series

$$\mathrm{Si}\, x = x - \frac{x^3}{3!\,3} + \frac{x^5}{5!\,5} - \ldots$$

3. To calculate $\sin x$ and $\cos x$ from this table in cases when their values are not given here the reduction formulae

$$\sin x = \sin (x - 2n\pi) ; \qquad \cos x = \cos (x - 2n\pi) ,$$

where $n = 1, 2, 3, \ldots$, must be used.

The values of $2n\pi$ are set out in the following table:

n	$2n\pi$	n	$2n\pi$
1	6·28	6	37·70
2	12·57	7	43·98
3	18·85	8	50·27
4	25·13	9	56·55
5	31·42	10	62·83

BIBLIOGRAPHY

(Containing sources and literature recommended for further study in the theory of automatic control)

General Courses on the Theory of Automatic Control

1. *Automatic Control* (Avtomaticheskoye regulierovaniye), a collection of material from the conference in Cranfield, 1951, translated from the English and edited by M. Z. Litvin-Sedy, Moscow, Inostran lit. (1954).
2. Z. Sh. Blokh, *The Control of Machines* (Regulirovaniye mashin) Moscow, Gostekhizdat (1950).
3. Z. Sh. Blokh, *The Dynamics of the Linear Systems used in the Automatic Control of Machines* (Dinamika lineinykh sistem avtomaticheskogo regulirovaniya mashin), Moscow, Gostekhizdat (1952).
4. G. Bode, *The Theory of Circuits and Design of Amplifiers with Feedback* (Teoriya tsepeie i proyektirovaniye usilitelei s obtratnoi svyaz'yu), Moscow, Inostran. lit. (1948).
5. A. A. Voronov, *The Elements of the Theory of Automatic Control* (Elementy teorii avtomaticheskogo regulirovaniya), Moscow, Voyenizdat (1954).
6. S. G. Gerasimov, *The Physical Basis of the Control of Thermal Processes* (Fizicheskiye osnovy reguliroveniya teplovykh protsessov), Moscow, Ob'ed. nauch. tekh. izd. (1937).
7. S. G. Gerasimov, *The Theoretical Basis of the Automatic Control of Thermal Processes* (Teoreticheskiye osnovy avtomaticheskogo regulirovaniya teplovykh protsessov), Moscow, Gosenergoizdat (1949).
8. S. G. Gerasimov, Ye. G. Dudnikov and S. F. Chistyakov, *The Automatic Control of Boiler Assemblies* (Avtomaticheskoye regulirovaniye kotel'nykh ustanovok), Moscow, Gosenergoizdat (1950).
9. Ye. G. Dudnikov, *The Essentials of the Automatic Control of Thermal Processes* (Osnovy avtomaticheskogo regulirovaniya teplovykh protsessov), Moscow, Gosenergoizdat (1956).
10. K. V. Yegorov, *The Essentials of Automatic Control* (Osnovy avtomaticheskogo regulirovaniya), Moscow, Gosenergoizdat (1955).
11. N. Ye. Zhukovskii, *The Theory of the Control of Machine Operation* (Teoriya regulirovaniya khoda mashin), Moscow, Gosmashmetizdat (1933).
12. A. G. Ivakhnenko, *Electroautomatics* (Elektroavtomatika) Parts I and II, Ukr. SSR, Gostekhizdat (1954).
13. A. G. Iosif'yan and B. M. Kogan, *The Essentials of Sequential Transmission* (Osnovy sledyashchego privoda), Moscow, Gosenergoizdat (1954).

14. A. A. KAMPE-NEMM, *The Dynamics of Two-position Control* (Dinamika dvukhpositsionnogo regulirovaniya), Moscow, Gosenergoizdat (1955).
15. S. A. KANTOR, *The Control of Turbines* (Regulirovaniye turbomashin), Moscow, Mashgiz (1946).
16. YU. G. KORNILOV and V. D. PIVEN', *The Essentials of the Theory of Automatic Control Applied to Thermal Energy Assemblies* (Osnovy teorii avtomaticheskogo regulirovaniya v prilozhenii k teplosilovym ustanovkam), Moscow, Mashgiz (1947).
17. N. T. KUZOVKOV, *The Theory of Automatic Control* (Teoriya avtomaticheskogo regulirovaniya), Moscow, Oborongiz (1957).
18. G. LAUER, S. LASTNIK and L. MADSON, *The Essentials of the Theory of Servomechanisms* (Osnovy teorii servomekhanizmov), Moscow, Gosenergoizdat (1948).
19. B. YA. LERNER, *Introduction to the Theory of Automatic Control* (Vvedeniye v teoriyu avtomaticheskogo regulirovaniya), Moscow, Mashgiz (1958).
20. V. L. LOSSIYEVSKII, *Automatic Controllers* (Avtomaticheskiye regulyatory), Moscow, Oborongiz (1944).
21. V. L. LOSSIYEVSKII, *The Essentials of the Automatic Control of Technological Processes* (Osnovy avtomaticheskogo regulirovaniya tekhnologicheskikh protsessov), Moscow, Oborongiz (1949).
22. L. MACCOLL, *The Essentials of the Theory of Servomechanisms* (Osnovy teorii servomekhanizmov), Moscow, Inostran. lit. (1947).
23. M. V. MEYEROV, *The Essentials of the Automatic Control of Electrical Machines* (Osnovy avtomaticheskogo regulirovaniya elektricheskykh mashin), Moscow, Gosenergoizdat (1952).
24. M. V. MEYEROV, *Introduction to the Dynamics of the Automatic Control of Electrical Machines* (Vvedeniye v dinamiku avtomaticheskogo regulirovaniya elektricheskykh mashin), Izd. akad. nauk SSSR (1956).
25. YE. L. NIKOLAI, *The Control of Machines* (Regulirovaniye mashin), Kubuch (1930).
26. R. OL'DENBURG and G. SARTORIUS, *The Dynamics of Automatic Control* (Dinamika avtomaticheskogo regulirovaniya), Moscow, Gosenergoizdat (1949).
27. V. V. SOLODOVNIKOV (Ed.), *The Essentials of Automatic Control* (Osnovy avtomaticheskogo regulirovaniya), Moscow, Mashgiz (1954).
28. YE. P. POPOV, *Automatic Control* (Avtomaticheskoye regulirovaniye), Moscow, Gostekhizdat (1956).
29. YE. P. POPOV, *The Dynamics of Systems of Automatic Control* (Dinamika sistem avtomaticheskogo regulirovaniya), Moscow, Gostekhizdat (1954).
30. *Sessions of the Academy of Sciences on the Scientific Problems of the Automation of Production* (Sessiya Akademii nauk po nauchnym problemam avtomatizatsii proizdvodstv), Izd. akad. nauk SSSR (1957).
31. J. JAMES, N. NICHOLS and R. PHILLIPS (Editors), *The Theory of Sequential Systems* (Teoriya sledyash'chikh sistem) translated under the editorship of Ya. Z. Tsypkin, Moscow, Inostran. lit. (1951).
32. M. TOLLE, *The Control of Motors* (Regulirovaniye dvigatelyei), Moscow, Gosenergoizdat (1950).

33. *The Works of the Second All-Union Conference on the Theory of Automatic Control* (Trudy Vtorogo vsesoyuznogo soveshchaniya po teorii avtomaticheskogo regulirovaniya), Vols. I, II and III, Izd. akad. nauk SSSR (1955—1956).
34. A. V. FATEYEV, *The Essentials of the Linear Theory of Automatic Control* (Osnovy lineinoi teorii avtomaticheskogo regulirovaniya), Moscow, Gosenergoizdat (1954).
35. A. A. FEL'DBAUM, *A Course of Electro-automatics* (Kurs elektroavtomatiki) (Part I — Drives and systems of electro-automatics; Part II — The elements of the dynamics of a system of automatic control), Moscow (1948).
36. A. A. FEL'DBAUM, *Electrical Systems of Automatic Control* (Elektricheskiye sistemy avtomaticheskogo regulirovaniya), Moscow, Oborongiz (1957).
37. TSYAN' SYUE-SEN, *Technical Cybernetics* (Tekhnicheskaya kibernetika), Inostran. lit. (1957).
38. *Special Methods in Automatics* (Chastotniye metody v avtomatike), collected articles, Inostran lit. (1957).
39. M. A. AIZERMAN, *Introduction to the Dynamics of the Automatic Control of Motors* (Vvedeniye v dinamiku avtomaticheskogo regulirovaniya digatelei), Moscow, Mashgiz (1950).
40. M. A. AIZERMAN, *The Theory of the Automatic Control of Motors* (Teoriya avtomaticheskogo regulirovaniya dvigatelei), Moscow, Gostekhizdat (1952).
41. W. AHRENDT and I. TAPLIN, *Automatic Regulation*, Vol. 1 (1948).
42. W. AHRENDT and I. TAPLIN, *Automatic Feedback Control*, McGraw-Hill (1951).
43. H. CHESTNUT and R. MAYER, *Servomechanisms and Regulating System* Design, Vol 1, New York—London (1951).
44. W. DUNCAN, *The Principles of the Control and Stability of Aircraft*, Cambridge University Press (1952).
45. W. R. EVANS, Control System Dynamics, New York-London (1955).
46. G. H. FARRINGTON, *Fundamentals of Automatic Control*, London, Chapman and Hall (1951).
47. A. HALL, *The Analysis and Synthesis of Linear Servomechanisms*, Massachusetts Institute of Technology (1947).
48. I. HANNY, *Regelungstheorie*, Kleines Handbuch für Ingenieure und Studierende, Zürich (1947).
49. A. LEONHARD, *Die selbstätige Regelung*, zweite Ausgabe, Berlin—Göttingen—Heidelberg (1957).
50. W. OPPELT, *Stetige Regelvorgänge* (1949).
51. W. OPPELT, *Grundsätze der Regelung*, Wolfenbüttel—Hannover (1947).
52. R. MACMILLAN, *An Introduction to the Theory of Control in Mechanical Engineering*, Cambridge University Press (1951).
53. P. PROFOS, *Vektorielle Regeltheorie*, Zürich (1944).
54. A. PORTER, *An Introduction to Servomechanisms*, Wiley (1950).
55. I. G. THALER and R. G. BROWN, *Servomechanisms Analysis*, Electrical and Electronic Engineering Series, McGraw-Hill (1953).
56. J. TRUXALL, *Automatic Feedback Control System Synthesis*, New York, McGraw-Hill (1955).

Chapter I

1. A. A. Andronov, I. A. *Vyshnegradskii and his role in creating the theory of automatic control*, Izv. akad. nauk SSSR, No. 5, p. 805 (1949).
2. A. A. Andronov and I. N. Voznesenskii, On the works of D. K. Maxwell, I. N. Vyshnegradskii and A. Stodola in the sphere of the theory of machine control, in the book: D. K. Maxwell, I. A. Vyshnegradskii and A. Stodola, *The Theory of Automatic Control (Linearizable Problems)* [Teoriya avtomaticheskogo regulirovaniya (linearizovanniye zadachi)]. Izd. Akad. Nauk SSSR (1949).
3. I. N. Voznesenskii, *On the principles and schemes of automatic control*, Prikl. mat. i mekh., Vol. VI, No. 1, p. 101 (1942). See also Trudy Leningr. politekhn. inta im. Kalinina, No. 2, p. 116 (1948).
4. G. S. Gorelik, *Radiophysics and the theory of automatic control*, Izv. akad. nauk SSSR, ser. fizich, No. 2, p. 103 (1947).
5. A. M. Kats, *New Designs for Precision Controllers of Imported Diesels* (Novyye konstruktsii pretsizionnykh regulyatorov importnykh dizelei), Nauch. issled. dizel. inst., Book 6 (Control and automatic on contemporary assemblies), Moscow, Mashgiz (1948).
6. A. M. Kats, *On the question of floating direct action controllers*, Automat. i Telemekh., No. 5 (1949).
7. V. L. Lossiyevskii, *Automatic Control* (Avtomaticheskoye regulirovaniye), Izd. Akad. Nauk SSSR (1946).
8. N. N. Nastenko, *All-state Control of Fast Moving Diesel Motors of the Transport Type* (Vserezhimnoyė regulirovaniye bistrokhodnykh dizelmotorov transportnovo tipa), Moscow, Mashgiz (1944).
9. A. Stodola, Inertial controllers (inertsionniye regulatory), in the book: D. Maxwell, I. A. Vyshnegrdskii and A. Stodola, *The Theory of Automatic Control* (Teoriya avtomaticheskogo regulirovaniya), in the series "Classics of Science" (Klassiki nauki), Izd. Akad. Nauk SSSR (1949).
10. V. A. Trapeznikov, *The principles of construction of industrial devices for automatic regulation and control*, Izv. akad. nauk SSSR, No. 10, p. 1450 (1950).
11. P. Tsemke, *Controllers of Piston Motors* (Regulyatory porshnevykh dvigatelei), Moscow, Gosmashmetizdat (1933).
12. V. M. Shlyandin, *The Elements of Automatics and Telemechanics* (Elementy avtomatiki i telemekhaniki), Moscow, Oborongiz (1954).
13. T. Shtein, *Control and Compensation in Steam Assemblies* (Regulirovaniye i vyravnivaniye v parovykh ustanovkakh), Moscow, GNTI (1931).
14. A. V. Shcheglyayev, *The Control of Steam Turbines* (Regulirovaniye parovykh turbin), Moscow, Gosenergoizdat (1933).
15. R. Proell, *Die Mechanik der Zentrifugeregulatoren und deren pseudoastatische Aufhängung*. Der Civilingenieur, p. 323 (1872).
16. W. Schmidt, *Unmittelbare Regelung*, Berlin (1944).

Chapter II

1. A. A. Abdulayev, I. V. Vaiser and E. M. Nadzhatov, *The equation of the pneumatic controller 04*, Automat. i Telemekh., No. 5 (1955).
2. S. G. Gerasimov (Ed.), *The Automatization of Thermal Processes* (Avtomatizatsiye teplovykh protsessov), Moscow, Gosenergoizdat (1948).
3. Z. Ya. Beirakh, *On the question of the character of the curve of non-uniformity of control*, Avtomat. i Telemekh., No. 6, p. 451 (1947).
4. I. N. Voznesenskii, On the question of the choice of a control system for throttle controlled turbines, in the Collection: *For Soviet Power Equipment* (Za sovyetskoye energooborudovaniye), Leningrad (1934).
5. I. A. Vyshnegradskii, On controllers of direct action, Izv. SPB Tekhnolog. Inst. (1877). (Reprinted in the book: *Theory of Automatic Control* (Teoriya avtomaticheskogo regulirovaniya), by D. Maxwell, I. A. Vyshnegradskii and A. Stodola, in the series "Classics of Science" (Klassiki nauki), lzd. Akad. Nauk SSSR (1949).
6. Ye. G. Dudnikov, l. M. Krassov, A. A. Tagayevskaya, V. P. Temnyi and P. T. Barkalov, *The experimental determination of the dynamic characteristics of industrial controlled objects*, Avtomat. i Telemekh., No. 4, p. 418 (1953).
7. L. A. Zalmanzon, *Graphs for determining the parameters of steady flow of air through systems of calibrated apertures in pneumatic controllers*, Avtomat. i Telemekh. ,No. 2, p. 176 (1952).
8. L. A. Zalmanzon, *On the differential equations of pressure change processes in flow chambers of pneumatic regulated and controlled assemblies*, Avtomat. i Telemekh., No. 3 (1954).
9. I. M. Krassov, A. A. Tagayevskaya and M. A. Vasilyeva, *The determination of the amplitude-phase characteristic of a controller, using the method of a rectangular wave*, Avtomat. i Telemekh., No. 3, p. 322 (1953).
10. V. L. Lossiyevskii, *The Application of the Theory of Similitude and Dynamic Analogues to the Problems of Models of Objects and Processes of Control* Primenyeniye trorii podobiya i dinamichesjih analogii k zadacham modelirovaniya obyektov. i protesesssov regulirovaniya), Moscow, Gosenergoizdat (1951).
11. D. Maxwell, On controllers (O regulyatorakh), in the book: *The Theory of Automatic Control* (Teoriya avtomaticheskogo regulirovaniya), by D. Maxwell, I. A. Vyshnegradskii and A. Stodola, in the series "Classics of Science" (Klassiki nauki), Izd. akad. nauk SSSR (1949).
12. M. V. Meyerov, *The investigation of a system for the control and regutation of the motor of a reversible rolling mill with metadyne controllers*, Elektrichestvo, No. 7, p. 15 (1949).
13. B. N. Petrov, *On the construction and transformation of structural networks*, Izv. akad. nauk SSSR, otdel. tekh. nauk, No. 12, p. 1146 (1945).
14. G. L. Rabkin, B. A. Mitrofanov and Yu. O. Shternberg, *On determining the numerical values of the coefficients of the transfer functions of linearizable stages and systems from their experimental frequency characteristics* A. and T., No. 5 (1955).

15. V. V. SOLODOVNIKOV, *The frequency response method in control theory* (Review), A. and T., No. 2 (1947).
16. V. V. SOLODOVNIKOV, *The application of the dynamic frequency response method to the investigation of stability and to the estimate of the quality of servomechanisms and controlled systems*, A. and T., No. 2, p. 85 (1948).
17. A. STODOLA, On the control of turbines (O regulirovanii turbin), in the book: *The Theory of Automatic Control* (Teoriya avtomaticheskogo regulirovaniya), by D. Maxwell, I. N. Vyshnegradskii and A. Stodola, in the series "Classics of Science" (Klassiki nauki), Izd. akad. nauk SSSR (1949).
18. A. A. TAGAYEVSKAYA, *The determination of the amplitude-phase characteristic of a linear system from the curve of its transient process*, Avtomat. i Telemekh., No. 2, p. 231 (1953).
19. M. A. AIZERMAN, *The dynamics of automatic pressure control in gas automobiles*, Nauchnoissledovatel'skii avtotraktornyi institut, Collection of scientific research works on autotraction, Moscow, Mashgiz (1938).
20. M. A. AIZERMAN, *The dynamics of the automatic multi-stage control of pressure*, Trudy NATI, Issue 38 (1940).
21. BEDFORD and FREDENDALL, *Analysis, Synthesis and Evaluation of the Transient Response of Television Apparatus*, Proc. IRE, Vol 30, p. 10 (1942).
22. I. HÄNNY, *Regelungtheorie*, Kleines Handbuch für Ingenieure und Studierende, Zürich (1949).
23. A. IVANOFF, *The Influence of the Characteristics of a Plant on the Performance of an Automatic Regulator*, Proc. Soc. of Chem. Ind., Vol. 18, p. 138 (1936).

Chapter III

1. A. A. ANDRONOV and A. G. MEYER, *The simplest linear systems with delays*, Avtomat. i Telemekh., Nos. 2 and 3 (1946).
2. G. V. ARONOVICH, *On the effect of an hydraulic shock on the stability of control of water turbines*, Avtomat. i Telemekh., No. 3 (1948).
3. T. T. BELOUSOVA, Diagrams of the Vyshnegradskii type for some systems of indirect control, in the collection: *The Control of Machines and the Synthesis of Mechanisms* (Regulirovaniye mashin i sintez mekhanizmov), edited by Prof. Z. Sh. Blokh, Moscow, Mashgiz (1950).
4. Z. SH. BLOKH, *A graphical analytic method for estimating the stability of linear control systems*, Avtomat. i Telemekh., No. 6, p. 441 (1947).
5. N. VOLOSHIN, *An account of the occurrence of delay*, Avtomat. i Telemekh., No. 4 (1948).
6. YU. V. VOROB'YEV, *On the investigation of the stability of a class of automatic control systems with wave processes in the separate stages*, Avtomat. i Telemekh., No. 4 (1948).
7. YU. V. VOROB'YEV, The stability of linear systems with wave processes in the separate stages, in the collection: *Research in the Subject of the*

Control of Steam Turbines (Issledovaniye v oblasti regulirovaniya parovykh turbin), Moscow, Gosenergoizdat, p. 128 (1950).

8. YU. V. VOROB'YEV and V. N. DROZDOVICH, *On methods for the investigation of the stability of control systems with distributed parameters*, Avtomat. i Telemekh., No. 27 (1949).
9. YU. V. VOROB'YEV, A. M. KATS and N. A. FOTEYEVA, The stability of a controller of direct action maintaining the pressure in a long pipe, in the Collection: *Research in the subject of the Control of Steam Turbines* (Issledovaniye v oblasti regulirovaniya parovykh turbin), Moscow, Gosenergoizdat, p. 175 (1950).
10. I. I. GALPERIN,* *Proof of the theorems concerning input substitution and the structural formula* (Dokazatel'stvo teorem o vkhodnom zameshchenii i o strukturnoi formule), Avtomat. i Telemekh., No. 3 (1951).
11. F. R. GANTMAKHER and M. G. KREIN, *Oscillating Matrices and Small Oscillations of Mechanical Systems* (Ostsillyatsionnye matritsy i malye kolebaniya mekhanicheskikh sistem), Moscow, Gostekhizdat (1950).
12. N. P. YERUGIN, *On a problem in the theory of the stability of an automatic control system*, Prikl. mat. i mekh., Vol. XVI, Issue 5 (1952).
13. N. P. YERUGIN, *On some questions of the stability of motion and qualitative theory of differential equations as a whole*, Prikl. mat. i mekh., Vol. XIV, Issue 5 (1950).
14. B. A. YERSHOV, *On the stability as a whole of a system of automatic control*, Prikl. mat. i mekh., Vol. XVII, Issue 1 (1953).
15. N. A. ZHELEZTSOV, *Vyshnegradskii diagrams for a floating controller of indirect action*, Avtomat. i Telemekh., No. 6 (1949).
16. *Notes from a seminar on the theory of the stability of motion*, Issues 1—4, Red Flag Military Air Engineering Academy (1947—1949).
17. I. P. KABAKOV and A. A. SOKOLOV, *On the effect of an hydraulic shock in a blocked oil pipe on the process of leverless flow control of the speed of steam turbines*, Inzh. sbornik, Vol. II, No. 2 (1946).
18. I. P. KABAKOV, *On the process of the control of steam pressure*, Inzh. sbornik, Vol. II, No. 2 (1946).
19. G. G. KALISH, *Research into the dynamics of a bosch-type centrifugal controller*, Dizelestroyeniye, No. 10 (1937).
20. G. G. KALISH and V. I. KRUTOV, A diagram of the stability of the process described by a linear differential equation of the third order, in the Collection: *The Stability of Operation of the Transport Diesel* (Ustoichivost raboty transportnovo dizelya), Moscow, Mashgiz (1949).
21. G. G. KALISH and V. I. KRUTOV, *The stability of the operation state of a transport diesel with variable speed control*, NAMI, Issue 51, The Stability of Operation of a Motor (Ustoichivost raboty dvigatelya), Moscow, Mashgiz (1948).
22. A. M. KATS and A. I. CHEKMAREV, *On the evaluation of a Hurwitz determinant*, Inzh. sbornik, Vol. XII (1952).

* For a discussi on of the works of I. I. Galperin and a list of his works, see *A. and T.*, No. 4 (1948), No. 1 (1949) and No. 3 (1951).

23. S. P. Kuvshinnikov, *The effect of a cataract on the dynamic stability of the indirect control of steam turbines*, Sovetskoye Kotloturbostroyeniye, No. 2 (1938).
24. A. M. Leetov, *The Stability of Non-linear systems of Automatic Control* (Ustoichivost nelineinykh sistem avtomaticheskogo regulirovaniya), Moscow, Gostekhizdat (1955).
25. A. I. Lur'ye and G. M. Fialko, *On the stability of control in the presence of a delay in the measuring element of the controller*, Inzh. sbornik, Vol. IV, No. 2 (1948).
26. A. M. Lyapunov, *A General Problem Concerning the Stability of Motion* (Obshchaya zadacha ob ustoichivosti dvizheniya), Ob'ed. nauch. tekh. izd. (1935).
27. I. G. Malkin, *On a problem in the theory of stability of an automatic control system*, Prikl. mat. i. mekh., Vol. XVI, Issue 3, pp. 365—368 (1952).
28. D. I. Maryanovskii, *Investigation of the stability of a control system when internal connexions are introduced*, Elektrichestvo, No. 3 (1950).
29. M. V. Meyerov, *Some questions of the choice of the structural scheme of a system of automatic control*, Avtomat. i Telemekh., No. 4, pp. 405—416 (1952).
30. M. V. Meyerov, *Some questions of the stability of the control of the output voltage of electric generators*, Vestnik elektropromyshlenosti, No. 9, pp. 11—15 (1943).
31. M. V. Meyerov, *On the use of the D-partition for the investigation of the quality of systems of automatic control*, Avtomat. i Telemekh., No. 6, p. 337 (1951).
32. M. V. Meyerov, *On a case of the use of the D-partition boundary for the evaluation of the quality of systems of automatic control*, Izv. akad. nauk SSSR, otdel. tekh. nauk, No. 12, p. 1784 (1950).
33. M. V. Meyerov, *On the calculation of small parameters for the investigation of the stability of systems of automatic control*, Elektrichestvo, No. 6 (1947).
34. M. V. Meyerov, On the choice of the parameters of systems of automatic control. In the book: *Proceedings of the Moscow Order of Lenin Energy Institute* (Trudy Moskovskogo ordena Lenina energetich. instituta) (1948).
35. M. V. Meyerov, On systems of auto-control, which are stable for any coefficient of amplification, however large, Avtomat. i Telemekh., No. 4 (1947).
36. M. V. Meyerov, *On the stabilization of systems containing elements with a delay*, Avtomat. i Telemekh., No. 5 (1953).
37. M. V. Meyerov, *The principles of construction of systems of auto-control with a small steady error*, Avtomat. i Telemekh., No. 2 (1949).
38. D. I. Maryanovskii, *A system of automatic control with rigid internal feedback*, Vestnik Elektropromyshlennosti, No. 3, p. 8 (1949).
39. D. I. Maryanovskii, *The stability of linear systems of automatic control*, Elektrichestvo, No. 9 (1946).
40. A. V. Mikhailov, *The method of harmonic analysis in the theory of control*, Avtomat. i Telemekh., No. 3 (1938).

41. A. V. Mikhailov, *The theory of the stability of linear feedback circuits with lumped constants*, Zh. Tekh., No. 1 (1939).
42. N. N. Myasnikov, *The Mikhailov criterion and the analysis of systems with a delay*, Proceedings of the Leningrad Red Flag Military Air Engineering Academy, Issue 36 (1951).
43. N. N. Myasnikov, *The Mikhailov criterion and the estimate of the roots of the characteristic equation*, Avtomat. i Telemekh., No. 4 (1949).
44. *On some cases of the distribution of the roots of the characteristic equation of a system of auto-control*, Proceedings of the Leningrad Red Flag M. A. E. A., Issue 36 (1951).
45. N. N. Myasnikov, *The theory of direct control according to Vyshnegradskii, and the influence of the delay effect*, Izv. akad. nauk SSSR, otdel. tekh. nauk, No. 9 (1953).
46. Yu. I. Neimark, *On the question of the effect of an hydraulic shock on the control of turbines*, Avtomat. i Telemekh., No. 4 (1948).
47. Yu. I. Neimark, *On the problem of the distribution of the roots of polynomials*, Dokl. akad. nauk SSSR, Vol. 58, No. 3 (1947).
48. Yu. I. Neimark, *On the determination of the values of the parameters for which a system of automatic control is stable*, Avtomat. i Telemekh., No. 3 (1948).
49. Yu. I. Neimark, *On the structure of the region of stability of single-loop systems*, Avtomat. i Telemekh., No. 1 (1950).
50. Yu. I. Neimark, *The structure of the D-partition of the space of polynomials and the Vyshnegradskii and Nyquist diagrams*, Dokl. akad. nauk SSSR, Vol. 59, No. 5 (1948).
51. Yu. I. Neimark, *The structure of the D-partition of the space of quasipolynomials and the Vyshneygradskii and Nyquist diagrams*, Dokl. akad. nauk SSSR, Vol. 50, No. 9 (1948).
52. Yu. I. Neimark, *The stability of linearizable systems (discrete and distributed)*, Leningrad Red Flag M. A. E. A. (1949).
53. V. D. Piven', *On floating control*, Sovietskoye Kotloturbostroyeniye, No. 10, p. 454 (1938).
54. L. S. Pontryagin, *On the zeros of some elementary transcendental functions*, Izv. akad. nauk SSSR, seriya matemat., Vol. 6, No. 3 (1942).
55. A. M. Popovskii, *On the construction of the D-partition for systems of control of interconnected quantities, given by experimentally determined characteristics*, Avtomat. i Telemekh., No. 3, p. 308 (1953).
56. A. M. Popovskii, *On the freedom of choice of the parameters of autonomous processes of control of several interconnected quantities*, Avtomat. i Telemekh., No. 6 (1949).
57. I. M. Sadovskii, *Nomograms for the checking of stability*, Avtomat. i Telemekh., No. 5 (1952).
58. A. A. Sokolov, *A criterion for the stability of linear systems of control with distributed parameters, and its application*, Inzh. sbornik, Vol. II, No. 2 (1946).
59. V. V. Solodovnikov, *The application of the operator method to the investiga-*

tion of the control process of the speed of a hydro-turbine, Avtomat. i Telemekh., No. 1 (1941).

60. V. V. Solodovnikov, Nyquist'scriterion for asymptotic stability. In the book: *Notes from the Seminar on the Theory of Stability of Motion* (Zapiski seminara po teorii ustoichivosti dvizheniya), Moscow, Leningrad Red Flag M. A. E. A. (1948).
61. V. V. Solodnikov, *The stability of linear servomechanisms and control systems*, Avtomat. i Telemekh., No. 1 (1946).
62. *The Works of the Second All-Union Conference on the Theory of Automatic Control* (Trudy vtorogo vsesoyuznogo soveshchaniya po teorii avtomaticheskogo regulirovaniya), Vol. I, Izd. akad. nauk SSSR (1955).
63. A. A. Fel'dbaum, *The connexion between statism and stability in an electric mechanism with a booster generator*, Vestnik Elektropromyshlennosti, No. 1 (1946).
64. Ya. Z. Tsypkin, A criterion for the stability of linear servomechanisms, in the book: *The Essentials of Servomechanism Theory* (Osnovy teorii servomekhanizmov) by G. Lauer, S. Lesnik and L. Madson, Moscow (1948).
65. Ya. Z. Tsypkin, *A criterion for the stability of linear systems with feedback*, Zh. tekh. fiz., No. 6 (1949).
66. Ya. Z. Tsypkin, *The stability of a class of automatic control system with distributed parameters*, Avtomat. i Telemekh., No. 3 (1948).
67. Ya. Z. Tsypkin, *The stability of systems with delayed feedback*, Avtomat i Telemekh., Nos. 2—3 (1946).
68. Ya. Z. Tsypkin, *The stability of systems with feedback*, Radiotekhnika, No. 5 (1946).
69. N. G. Chebotarev, *The Hurwitz problem for transcendental functions*, Works of the Leningrad Red Flag M. A. E. A., Issue 7 (1945).
70. N. G. Chebotarev, *On the Hurwitz problem for integral transcendental functions*, Dokl. akad. nauk SSSR, nov. ser., Vol. XXXIII, No. 9 (1941).
71. N. G. Chebotarev, *On a modification of the Sturm and Fourier methods*, Dokl. akad. nauk SSSR, nov. ser, Vol. XXXIV, No. 1 (1942).
72. N. G. Chebotarev, *On a modified statement of the Hurwitz problem*, Dokl. akad. nauk SSSR, nov. serv., Vol. XXXIV, No. 8 (1942).
73. N. G. Chebotarev, *On a special form of transcendental equation*, Dokl. akad. nauk SSSR, nov. serv., Vol. XXXIV, No. 2 (1942).
74. N. G. Chebotarev, *On integral functions with real alterntaing roots*, Dokl. akad. nauk SSSR, nov. ser., Vol. XXXV, No. 7 (1942).
75. N. G. Chebotarev and N. N. Meiman, *The Routh—Hurwitz problem of stability for polynomials and integral functions*, Trudy Mat. Inst. Im. V. A. Steklova, Vol. XXVI (1949).
76. N. G. Chetayev, *The stability of motion* (Ustoichivost dvizheniya), Moscow, Gostekhizdat (1955). English translation, Pergamon Press (1961).
77. M. A. Aizerman, *Sufficient conditions for the stability of a class of dynamic systems with variable parameters*, Prikl. mat. i mekh., Vol. XV, Issue 3 (1951).
78. M. A. Aizerman, *A problem concerning the stability of the process of direct*

control of the speed of a motor when the non-linearities of its characteristic are taken into account, Trudy NAMI, Issue 51, The Stability of Operation of a Motor (Ustoichivost' raboty dvigatelya), Moscow, Mashgiz (1948).

79. M. A. AIZERMAN, *Regions of Stability and their Isolation* (Oblasti nastoichivosti i ikh vydeleniye). Lectures to the Seminar on the Theory of Automatic Control, Moscow, Mashgiz (1950).
80. M. A. AIZERMAN, *On a problem concerning the "large" stability of dynamic systems*, Usp. mat. nauk, Vol. IV p. 186 (1949).
81. M. A. AIZERMAN, *On increasing the critical value of the coefficient of amplification of a single-loop system of automatic control by the introduction of derivative action*, Avtomat. i Telemekh., No. 2 (1951).
82. M. A. AIZERMAN, *Taking non-linear functions of several arguments into account in the investigation of the stability of systems of automatic control*, Avtomat. i Telemekh., No. 1 (1947).
83. M. A. AIZERMAN, *On some structural conditions of stability for systems of automatic control*, Avtomat. i Telemekh., No. 2 (1948).
84. M. A. AIZERMAN, *On the practical application of the Hurwitz criterion*, Avtomat. i Telemekh., No. 2 (1952).
85. M. A. AIZERMAN, *On the convergence of the control process after large initial deflections*, Avtomat. i Telemekh., Nos. 2—3 (1946).
86. M. A. AIZERMAN, and F. R. GANTMAKHER, *Conditions for the existence of the region of stability for a single-loop system of automatic control*, Prikl. mat. i mekh., Vol. XVIII, No. 1 (1954).
87. A. HURWITZ, *Über die Bedingungen unter welcher eine Gleichung nur Wurzeln mit negativen reellen Teilen besitzt*, Mathem. Ann., Vol. XLVI (1895).
88. K. KÜPFMÜLLER, *Über die Dynamik der selbststätigen Verstärkungsregler*, El. Nachrichten Technik, Vol. 5, p. 459 (1928).
89. PRINZ, *Contributions to the Theory of Automatic Controllers and Followers*, Journ. Scient. Instr., Vol. IV (1944).

Chapter IV

1. D. A. BASHKIROV, *A Graphic-Analytic Method for the Anlaysis of Transient Processes in Systems of Automatic Control* (Grafoanaliticheskii metod analiza perekhodnykh protsessov v sistemakh avtomaticheskogo regulirovaniya), Candidate dissertation, Leningrad (1950).
2. D. A. BASHKIROV, *A method for the construction of transient processes in non-linear systems of automatic control*, Proceedings of the Second All-Union Conference on the Theory of Automatic Control, Vol. II, Izd. akad. nauk SSSR (1955).
3. G. A. BENDRIKOV and K. F. TEODORCHIK, *Laws of the migration of the roots of linear algebraic equations of the third and fourth degrees, with a continuous change in the free term*, Avtomat. i Telemekh., No. 3 (1955).
4. Z. SH. BLOKH, *The analysis of the quality of control in single-loop networks*,

in the collection: Units and Elements of the Theory of Automatics and Telemechanics (Ustroistva i elementy teorii avtomatiki i telemekhaniki), Moscow, Mashgiz (1952).

5. Z. SH. BLOKH, *On the analysis of the quality of transient processes in systems of automatic control.* In the collection: The Control of Machines and the Synthesis of Mechanisms (Regulirovanize mashin i sintez mekhanismov), Moscow, Mashgiz (1950).
6. Z. SH. BLOKH, *On the determination of the control time*, in the collection: Units and Elements of the Theory of Automatics and Telemechanics (Ustroistva i elementy teorii avtomatiki i telemekhaniki), Moscow, Mashgiz (1952).
7. Z. SH. BLOKH, *Some estimates of the quality of control, using frequency characteristics*, Avtomat. i Telemekh., No. 3 (1955).
8. Z. SH. BLOKH, *On defining the boundaries of monotonicity of transient processes more exactly.* In the collection: Units and Elements of the Theory of Automatics and Telemechanics (Ustroistva i elementy teorii avtomatiki i telemekhaniki), Moscow, Mashgiz (1952).
9. Z. SH. BLOKH, *On the monotonicity of transient processes in linear systems of automatic control*, No. 2, Avtaomat. i Telemekh., No. 2 (1950).
10. V. A. BODNER, *On the Choice of Optimal Parameters of Controlled Systems* (O vyborye optimal'nykh parametrov reguliruyemykh sistem), Moscow, Oborongiz (1953).
11. B. V. BULGAKOV, *On the accumulation of disturbances in linear oscillatory systems with constant parameters*, Dokl. akad. nauk SSSR, nov. ser., Vol. 51, No. 5, p. 339 (1946).
12. I. V. VAISER, *A remark on the use of the concept "degree of stability" for estimating a control process*, Avtomat. i Telemekh., No. 5 (1955).
13. I. N. VOZNESENSKII, *On the control of machines with a large number of controlled parameters*, Trudy Leningradskogo politekhn. in-ta im. Kalinina, No. 2 (1948). See also Avtomat. i Telemekh., Nos. 4—5 (1938).
14. A. A. VORONOV, *On the approximate construction of curves of the transient process from the real frequency characteristic*, Avtomat. i Telemekh., No. 6 (1952).
15. L. S. GOL'DFARB, *The conditions for aperiodicity of some systems of autocontrol*, byul. VEI, No. 8 (1940).
16. A. M. KATS, *On the question of the calculation of the quadratic criterion for the quality of control*, Prik. mat. i mekh., Vol. XVI, Issue 3 (1952).
17. A. M. KATS, *Some Questions of the Calculation of Controllers of Direct Action for Diesels* (Nyekotorye voprosy rascheta regulyatorov pryamogo deistviya dlya dizelei), NIDI, Book 10 (Questions of the Design of diesels), Moscow, Mashgiz (1949).
18. A. M. KATS, *The determination of the parameters of a controller from the given characteristic equation of the control system*, Avtomat. i Telemekh., No. 3 (1955).
19. A. M. KATS, *A comparative analysis of the dynamics of the most important schemes of indirect control of the speed of diesels*, in the Collection of the Scientific-Research Diesel Institute, Issue 16: Questions of the Design

of the Auxiliary Systems of Diesels (Voprosy konstruirovaniya vspomogatel'nykh agregator dizelei), Moscow, Mashgiz (1950).

20. M. Z. LITVIN-SEDYI (Ed.), *Correcting Circuits in Automatics* (Korrektiruyushchiye tsepi v avtomatikye), A collection of translations of articles, IL, Moscow (1954).
21. A. A. KRASOVSKII and G. S. POSPELOV, *Some methods for the calculation of approximate time characteristics of linear systems of automatic control*, Avtomat. i Telemekh., No. 6 (1953).
22. O. M. KRYZHANOVSKII, *On an iterative method for determining approximate values for the roots of equations*, Avtomat. i Telemekh., No. 5 (1950).
23. O. M. KRYZHANOVSKII, *The calculation of integral criteria for optimal control processes*, Dokl. akad. nauk Ukr. SSR, No. 3 (1953).
24. O. M. KRYZHANOVSKII, *On Quadratic Criteria for the Quality of Transient Processes of Control*, Described by Linear Difference Equations with Constant Coefficients, Dokl. akad. nauk Ukr. SSR, No. 3 (1953).
25. N. T. KUZOVKOV, *The use of the method of logarithmic frequency characteristics for the estimate of damping in control systems*, Avtomat. i Telemekh., No. 4 (1954).
26. N. B. KUTIN, *On the computation of the correlation function of a stationary random process from experimental data*, Avtomat. i Telemekh., No. 3 (1957).
27. V. I. KUKHTENKO, *On the combined reproduction of random and unrandom signals by linear servomechanisms*, Avtomat. i Telemekh., No. 5 (1957).
28. D. I. MARYANOVSKII, *Changing the parameters of control systems by internal connexions*, Elektrichestvo, No. 2 (1950).
29. M. V. MEYEROV, *A criterion for the aperiodicity of control*, Izv. akad. nauk SSSR, Otdel. tekh. nauk, No. 12 (1945).
30. A. V. MIKHAILOV, *A criterion for the aperiodicity of auto-controlled systems*, Avtomat. i Telemekh., No. 1 (1941).
31. V. S. PUGACHEV, *The theory of random functions and its application to problems of automatic control*, (Teoriya sluchainykh funktsii i yeyo primeneniye k zadacham avtomaticheskovo upravleniya), Gostekhizdat (1957).
32. A. M. RUBINCHIK, *A graphical method for the determination of the region of distribution of the roots of characteristic equations*, in the collection: The Control of Machines and the Synthesis of Mechanisms (Regulirovanye mashin i sintez mekhanismov), Moscow, Mashgiz (1950).
33. A. M. RUBINCHIK, *On the selection of the parameters for stability of a system of automatic control*, in the collection: The Control of Machines and the Synthesis of Mechanisms (Regulirovaniye mashin i sintez mekhanismov), Moscow, Mashgiz (1950).
34. A. M. RUBINCHIK, *An approximate method for estimating the quality of control in linear systems*, in the collection: The Units and Elements of the Theory of Automatics and Telemechanics (Ustroistva i elementy teorii avtomatiki i telemekhaniki), Moscow, Mashgiz (1952).
35. V. V. SOLODOVNIKOV, *The analysis of the quality of servomechanisms from their amplitude-phase characteristics*, Izv. akad. nauk SSSR, Otdel. tekh nauk, No. 4, p. 473 (1949).

36. V. V. SOLODOVNIKOV, *Introduction to the statistical dynamics of systems of automatic control* (Vvredeniye v statisticheskuye dinamiku sistem avtomaticheskovo upravleniya), Moscow, Gostekhizdat (1952).
37. V. V. SOLODOVNIKOV, *The investigation of the dynamics of electric servomechanisms and systems of auto-control by the frequency/characteristic method*, Elektrichestvo, No. 4 (1947).
38. V. V. SOLODOVNIKOV, *A criterion for the quality of control*, Dokl. akad. nauk SSSR, nov. ser., Vol. 60, p. 977 (1948).
39. V. V. SOLODOVNIKOV, *A criterion for the absence of overshoot and a criterion for monotonicity*, Dokl. akad. nauk SSSR, nov. ser., Vol. 62 p. 599 (1948).
40. V. V. SOLODOVNIKOV, *On an approximate method for the investigation of the dynamics of control systems and of servomechanisms*, Izv. akad. nauk SSSR, Otdel. tekh. nauk, No. 12, p. 1179 (1945).
41. V. V. SOLODOVNIKOV, *On the control equation and on a basic problem in control theory*, by VEI, No. 4 (1941).
42. V. V. SOLODOVNIKOV, *On the conditions under which overshoot cannot be large*, Avtomat. i Telemekh., No. 4 (1948).
43. V. V. SOLODOVNIKOV, *On the dynamic precision and optimal characteristics of servomechanisms and transducing devices*, Dokl. akad. nauk SSSR, nov. serv., Vol. 77, No. 2, p. 269 (1951).
44. V. V. SOLODOVNIKOV, *On the application of trapezoidal frequency characteristics to the analysis of the quality of automatic control systems*, Avtomat. i Telemekh., No. 5 (1949).
45. V. V. SOLODOVNIKOV, *A synthesis of correction feedback devices for servomechanisms with typical disturbances*, Avtomat. i Telemekh., No. 5 (1951).
46. V. V. SOLODOVNIKOV, *A synthesis of correction feedback devices for servomechanisms using optimal and typical logarithmic frequency characteristics*, Avtomat. i Telemekh., No. 5 (1953).
47. V. V. SOLODOVNIKOV, YU. I. TOPCHIYEV and G. V. KRUTIKOV, *The Frequency Method of Construction of Transient Processes with the Use of Tables and Nomograms* (Chastotnyi metod postroyeniya perekhodnykh protsessov s prilozheniem tablits i nomogramm), Moscow, Gostekhizdat (1955).
48. *Works of the Second All-Union Conference on the Theory of Automatic Control* (Trudy Vtorogo vsesoyuznogo soveshchaniya po teorii avtomaticheskogo reguliroraniya), Vol. II, Izd. akad. nauk SSSR, Moscow (1955).
49. A. I. TUPITSYN, *On the shape of the transient process curve for a minimum of integral error squared*, Avtomat. i Telemekh., No. 4 (1953).
50. G. M. ULANOV, *On the maximal deviation of the controlled quantity in the transient process*, Avtomat. i Telemekh., No. 3 (1948).
51. A. A. FEL'DBAUM, *Integral criteria for the quality of control*, Avtomat. i Telemekh., No. 1 (1948).
52. A. A. FEL'DBAUM, *Investigation of the dynamics of automatic control systems using the generalized integral criterion method*, Elektrichestvo, No. 7 (1951).
53. A. A. FEL'DBAUM, *On the distribution of the roots of the characteristic equations of control systems*, Avtomat. i Telemekh., No. 4 (1948).

54. YA. Z. TSYPKIN, *On the upper bound of the degree of stability of single-loop systems of automatic control*, Avtomat. i Telemekh., No. 4 (1952).
55. YA. Z. TSYPKIN, *The degree of stability of systems with delayed feedback*, Avtomat. i Telemekh., No. 3 (1947).
56. YA. Z. TSYPKIN and P. V. BROMBERG, *On the degree of stability of linear systems*, Izv. akad. nauk SSSR, Otdel. tekh. nauk, No. 12 (1945).
57. YA. Z. TSYPKIN and P. V. BROMBERG, *The degree of stability of linear systems*, Trudy NISO, Izd. BNT, MAP, No. 5 (1946).
58. M. A. AIZERMAN, *On the damping of oscillatory motion, characterized by a linear differential equation of the third order with constant coefficients*, Automat. i Telemekh., No. 1 (1940).
59. W. R. EVANS, *Control System Synthesis by Root Locus Method*, Trans. AIEE, Vol. 69 p. 66 (1950) ; Electr. Engineering No. 5 p. 405 (1950).
60. A. C. HALL, *Application of Circuit Theory to Design of Servomechanisms*, Journ. Frankl. Inst., Vol. 4, p. 279 (1946).
61. D. W. RUSSELL and C. H. WEAVER, *Synthesis of Servomechanisms by Root Locations*, Electrical Engineering, Vol. 72, p. 41 (1953) ; Trans. AIEE, Vol. 71, pp. 95—104 (1952).
62. H. A. SAMULON, *Spectrum Analysis of Transient Response Curves*, Proc. IRE, Vol. 39, No. 2 (1951).

Chapter V

1. A. A. ANDRONOV and E. S. CHAIKIN, *The Theory of Oscillations* (Teoriya kolyebanii), Ob'ed. nauch. tekh. inst. (1937).
2. A. A. ANDRONOV and N. N. BAUTIN, *The motion of a neutral aeroplane, supplied with an autopilot, and the theory of the point transformations of surfaces*, Dokl. akad. nauk. SSSR, nov. ser., Vol. 43, No. 5 (1945).
3. A. A. ANDRONOV and N. N. BAUTIN, *On a degenerate case of the general problem of direct control*, Dokl. akad. nauk SSSR, nov. ser., Vol. 46, No. 7 (1945).
4. A. A. ANDRONOV and N. N. BAUTIN, *The stabilization of the course of a neutral aeroplane by an autopilot with constant velocity of the servomotor and zone of insensitivity*, Dokl. akad. nauk SSSR, nov. ser., Vol. 46, No. 4 (1945).
5. A. A. ANDRONOV, N. N. BAUTIN and G. S. GORELIK, *Auto-oscillations of the simplest circuit, containing a ship's screw automatically controlled in accordance with measured error*, Dokl. akad. nauk SSSR, nov. ser., Vol. 47, No. 4 (1945).
6. A. A. ANDRONOV, N. N. BAUTIN and G. S. GORELIK, *The theory of indirect control when the coulomb friction in the sensor is taken into account*, Automat i Telemekh., No. 1 (1946).
7. A. A. ANDRONOV and A. G. MAIYER, *The Vyshnegradskii problem in the theory of direct control*, Avtomat. i Telemekh., No. 5 (1947).
8. A. A. ANDRONOV and A. G. MAIYER, *The Vyshnegradskii problem in the theory of direct control, II (the theory of of the direct action controller in*

the presence of coulomb and viscous friction, Avtomat. i Telemekh., No. 5 (1953).

9. A. A. Andronov and A. G. Maiyer, *The Mises problem in the theory of direct control and the theory of the point transformation of surfaces*, Dokl. akad. nauk SSSR, nov. ser., Vol. 43 (1944).
10. A. A. Andronov and A. G. Maiyer, *On the Vyshnegradskii problem in the theory of direct control*, Dokl. akad. nauk SSSR, nov. ser., Vol. 47, No. 5 (1945).
11. N. N. Bautin, *The Behaviour of Dynamic Systems near the Boundary of the Region of Stability* (Povedeniye dinamicheskikh sistem vblizi granits oblasti ustoichivosti), Moscow, Gostekhizdat (1949).
12. V. A. Besekerskil, *The use of vibrators for the elimination of non-linearities in automatic controllers*, Avtomat. i Telemekh., No. 6 (1947).
13. P. V. Bromberg, *Stability and Auto-oscillations of Impulse Systems of Control* (Ustoichivost i avtokolyebaniya impul' snykh sistem regulirovaniya), Moscow, Oborongiz (1953).
14. B. V. Bulgakov, *Auto-oscillations of control systems*, Dokl. akad. nauk SSSR, nov. ser., Vol. 37, No. 9, p. 283 (1942).
15. B. V. Bulgakov, *Auto-oscillations of control systems*, Prikl. mat. i mekh., VII, No. 2 (1945).
16. B. V. Bulgakov, *Oscillations* (Kolyebaniya), Moscow, Gostekhizdat (1954).
17. B. V. Bulgakov, *Some Problems of the Theory of Control with Non-Linear Characteristics*, Prikl. mat. i mekh., Vol. X, No. 3 (1946).
18. B. V. Bulgakov, *On the application of the Van der Pol method to pseudolinear oscillatory systems with many degrees of freedom*, Prikl. mat. i mekh., Vol. VI., No. 6 (1942).
19. N. V. Butenin, *The action of external sinusoidal forces on an auto-oscillatory system with gyroscopic forces*, Works of the Leningrad Red Flag M. A. E. A., Issue 7 (1945).
20. N. V. Butenin, *Mechanical Auto-oscillatory Systems with Gyroscopic Forces*, Prikl. mat. i mekh., Vol. IV, No. 5 (1942).
21. A. A. Voronov, *The approximate determination of the process of establishment of auto-oscillations in some non-linear systems of automatic control*, Avtomat. i Telemekh., No. 6 (1957).
22. L. S. Gol'dfarb, *On some non-linearities in control systems*, Avtomat. i Telemekh., No. 5 (1947).
23. L. S. Gol'dfarb, *On the non-linearity of control systems*, VEI, No. 3 (1941).
24. G. S. Gorelik, *Oscillations and Waves* (Kolyebaniya i volny), Moscow, Gostekhizdat (1950).
25. Ya. I. Grdina, *The motion of a direct action controller and its stability*, Vestnik Obshchestva Tekhnologov, Nos. 4—7 (1898).
26. Ya. I. Grdina, *On the question of the dynamic stability of centrifugal controllers*, Izvestiya Dnepropetrovskogo instituta im. Artema Sergeyeva, p. 41 (May 1927).
27. Yu. V. Dolgolenko, *The stability and auto-oscillations of a relay control system with a delay*, Avtomat. i Telemekh., No. 2 (1952).
28. Yu. V. Dolgolenko, *Fluttering states in relay systems of indirect control*,

Works of the Second All-Union Conference on the Theory of Automatic Control, Vol. 1, Izd. akad. nauk SSSR (1955).

29. Yu. V. Dolgolenko, *The approximate determination of partially-fluttering periodic states in relay systems of control*, Avtomat. i Telemekh., No. 1 (1957).
30. A. P. Duvakin and A. M. Letov, *On the stability of controlled systems with two regulating units*. Prikl. mat. i. mekh., Vol. XVIII, No. 2 (1954).
31. Ye. G. Dudnikov, *Introduction to the theory of continuous control with a constant closing velocity*, Avtomat. i Telemekh. No. 6 (1939).
32. L. A. Zalmanzon, *On taking the effect of the non-linearity of chambers of pneumatic controllers into account in control processes*, Avtomat. i Telemekh., No. 5 (1955).
33. B. M. Kagan, *The frequency method of construction and investigation of transient processes in non-linear systems of automatic control*, Works of the Second All-Union Conference on the Theory of Automatic Control, Vol. II, Izd. akad. nauk SSSR (1955).
34. V. V. Kazakevich, *On the auto-oscillations propagated in control systems by falling characteristics of friction in servomotors*, Avtomat. i Telemekh., No. 6 (1951).
35. A. M. Kats, The influence of Coulomb friction on the process of indirect control, in the collection: *Research in the Sphere of the Control of Steam Turbines* (Issledovaniya v oblasti regulirovaniya parovykh turbin), Moscow, Gosenergoizdat (1950).
36. Yu. G. Kornilov, *On the effect of insensitivity on the dynamics of indirect control*, Avtomat. i Telemekh., No. 1 (1950).
37. Yu. G. Kornilov, *Controllers with servomotors of constant velocity in a "fluttering state"*, Avtomaticheskoye regulirovaniye, Issue 2, Moscow, Mashgiz (1951).
38. Yu. G. Kornilov, *The theory of fluttering states in relay systems of automatic control*, Works of the Second All-Union Conference on the Theory of Automatic Control, Vol. I, Izd. akad. nauk SSSR (1955).
39. N. A. Korolev, *On periodic states in relay systems with internal feedback*, Avtomat. i Telemekh., No. 11 (1956).
40. A. A. Krasovskii, *On the vibration method of linearization of some non-linear systems*, Avtomat. i Telemekh., No. 1 (1948).
41. A. A. Krasovskii, *An estimate of the deflections in the simplest relay systems of automatic control*, Avtomat. i Telemekh., No. 2 (1953).
42. N. M. Krylov and N. N. Bogolyubov, *Introduction to Non-Linear Mechanics* (Vvedeniye v nelineinuyu mekhaniku), Izd. akad. nauk USSR (1937).
43. A. Ya. Lerner, *On the high-speed limiting of systems of automatic control*, Avtomat. i Telemekh., No. 6 (1954).
44. A. Ya. Lerner, *Improvement of the dynamic properties of automatic compensators with the help of non-linear elements, I*, Avtomat. i Telemekh., No. 2 (1952).
45. A. Ya. Lerner, *Improvement of the dynamic properties of automatic compensators with the help of non-linear elements, II*, Avtomat. i Telemekh., No. 4 (1952).

46. A. M. LETOV, *Boundary values of the smallest characteristic number of a class of controlled system*, Prikl. mat. i mekh., Vol. XV, No. 5, p. 591 (1951).
47. A. M. LETOV, *On the problem of the autopilot*, Vestn. Mosk. universiteta, No. 1, p. 123 (1946).
48. A. M. LETOV, *On the theory of floating controllers*, Prikl. mat. i mekh., Vol. XII, No. 4 (1948).
49. A. M. LETOV, *On a particular investigation of the stability of control systems*, Prikl. mat. i mekh., Vol. XII, No. 6 (1948).
50. A. M. LETOV, *Control of the stationary state of a system, maintained by the action of constant disturbing forces*, Prikl. mat. i mekh., Vol. XII, No. 2 (1948).
51. A. M. LETOV, *Especially unstable controlled systems*, Prikl. mat. i mekh., Vol. XIV, No. 2 (1950).
52. A. M. LETOV, *The stability of control systems with two regulating units*, Prikl. mat. i mekh., Vol. XIV, No. 4 (1953).
53. A. M. LETOV, *The Stability of Non-Linear Systems of Automatic Control* (Ustoichivost nelineinykh sistem avtomaticheskogo regulirovaniya), Moscow, Gostekhizdat (1955).
54. L. YA. LEKHTMAN, *On the analysis of relay servomechanisms*, Automat. i Telemekh., No. 1 (1951).
55. A. I. LUR'YE, *The effect of the force of friction in the measuring element of a controller on the process of indirect control*, Inzh. sbornik, Vol. IV, No. 1 (1947).
56. A. I. LUR'YE, *Some non-linear problems of the theory of automatic control* (Nyekotorye nelineinye zadachi teorii avtomaticheskogo regulirovaniya) Moscow, Gostekhizdat (1951).
57. A. I. LUR'YE, *On auto-oscillations in some controlled systems*, Avtomat. i Telemekh., No. 5 (1947).
58. A. I. LUR'YE, *On the stability of motion of a dynamic system*, Prikl. mat. i mekh., Vol. XI, No. 4 (1947).
59. A. I. LUR'YE, *On the stability of a certain class of controlled systems*, Prikl. mat. i mekh., Vol. IX, No. 5 (1945).
60. A. I. LUR'YE, *On the canonical form of the equations of the theory of automatic control*, Prikl. mat. i mekh., Vol. XII, No. 5 (1948).
61. A. I. LUR'YE, *On the character of the boundaries of the region of stability of controlled systems*, Prikl. mat. i mekh., Vol. XIV, No. 4 (1950).
62. A. I. LUR'YE and V. N. POSTNIKOV, *On the theory of stability of controlled systems*, Prikl. mat. i mekh., Vol. VIII, No. 3 (1944).
63. I. G. MALKIN, *The Lyapunov and Poincaré methods in the theory of non-linear oscillations* (Metody Lyapunova i Puankare v teorii nelineinykh kolyebanii), Moscow, Gostekhizdat (1949).
64. V. D. MIRONOV, *On the Calculation of Floating Controllers with Constant Velocity*, Dissertation for the Degree of Candidate of Technical Sciences (K raschetu izodromnykh regulyatorov s postoyannoi skorost'yu). Vses. teplotekh. in-t im. Dzerzhinskogo (1947).
65. E. M. NADZHAFOV, *The approximate determination of the periodic solutions*

in systems of automatic control containing several non-linearities, Works of the Second All-Union Conference on the Theory of Automatic Control, Vol. I, Izd. akad. nauk SSSR (1955).

66. B. N. NAUMOV, *The effect of derivative action on the frequency and amplitude of the auto-oscillations in systems of control*, Collection of Works on Automatics and Telemechanics, Moscow, Izd. akad. nauk SSSR (1953).
67. YU. I. NEIMARK, *On the periodic motion of relay systems*, Collection in Memory of Acad. A. A. Andronov, Izd. akad. nauk SSSR (1955).
68. YU. I. NEIMARK, *On the periodic states and stability of relay systems*, Avtomat. i Telemekh., No. 5 (1953).
69. YU. I. NAIMARK, *On the auto-oscillations and forced oscillations of relay systems with a delay*, Automat. i Telemekh., No. 3 (1955).
70. YU. I. NEIMARK, *On fluttering states and periodic motions of a relay system*, Works of the Gorky Research Physico-Technological Institute, Gorkii (1956).
71. YU. I. NEIMARK, *On fluttering states in relay systems of automatic control*, Automat. i Telemekh., No. 1 (1957).
72. YU. I. NEIMARK and I. M. KUBLANOV, *Investigation of the periodic states and their stability for the simplest distributed system of relay control of temperature*, Automat. i Telemekh., No. 1 (1953).
73. YE. L. NIKOLAI, *On the calculation of a centrifugal controller of direct action*, Jubilee Collection of the Scientific-Technical Group of the Leningrad Technological Institute (1928).
74. G. N. NIKOL'SKII, *On the question of the automatic stability of boats on a set course*, Works of the Central Laboratory of Wired Connexions, Issue 1 (1934).
75. G. N. NIKOL'SKII, *On a problem of indirect control*, Inzh. sbornik, Vol. IV, No. 2 (1948).
76. V. V. PETROV, *On the auto-oscillations of servomechanisms with two-stage relay control*, Automat. i Telemekh., No. 1 (1951).
77. V. V. PETROV and G. M. ULANOV, *The use of proportional and emergency feedback for the suppression of the auto-oscillations in a servomechanism with two-stage relay control*, Avtomat. i Telemekh., No. 2 (1952).
78. YE. P. POPOV, *On the small parameter in the method of harmonic linearization* Izd. akad. nauk SSSR, Otdel. tekh. nauk, No. 2 (1955).
79. YE. P. POPOV, *Approximate investigation of the auto-oscillations and forced oscillations of high order non-linear systems on the basis of the harmonic linearization of the non-linearities*, Izv. akad. nauk SSSR, No. 5 (1954).
80. YE. P. POPOV, *An account of the effect of non-linearities in the calculation of servomechanisms*, Avtomat. i Telemekh., No. 6 (1953).
81. A. M. POPOVSKII, *The linearization of the fluttering state of operation of an on-off controller employing a constant speed servomotor*, Automat. i Telemekh., No. 3 (1950).
82. K. E. RERIKH, *The effect of friction on the process of control by centrifugal controllers of direct action*, Izvestiya Yekaterinoslavskogo gornogo instituta, XIV, Jubilee Issue (1899—1924), Dnepropetrovsk (1928).

83. B. A. Ryabov, *The determination of the parameters of the state of steady auto-oscillations of some systems*, Dokl. akad. nauk SSSR, nov. ser., Vol. 71, No. 4, p. 663 (1950).
84. I. M. Smirnova, *On the approximate investigation of the conditions of stability of periodic states in systems of automatic control*, Avtomat. i Telemekh., No. 2 (1954).
85. A. A. Sokolov, *The effect of insensitivity on the process of indirect control of the speed of steam turbines, in the collections:* Research in the Sphere of the Control of Steam Turbines (Issledovaniya v oblasti regulirovaniya parovykh turbin), Moscow, Gosenergoizdat (1950).
86. S. P. Strelkov, *Introduction to the Theory of Oscillations* (Vvedeniye v teorii kolyebanii), Moscow, Gostekhizdat (1950).
87. K. P. Surovikhin, *On the effect of the coefficient of amplification of feedback on the frequency and amplitude of auto-oscillations*, Avtomat. i Telemekh., No. 5 (1955).
88. A. A. Tal', *The Vyshnegradskii problem in the theory of direct control* (with an account of self-regulation of the object and derivative action of the controlled parameter), Avtomat. i Telemekh., No. 5 (1953).
89. K. F. Teodorchik, *Auto-oscillatory systems* (Avtokolyebatel'nye sistemy), Moscow, Gostekhizdat (1948).
90. K. F. Teodorchik, *Types of motion of relay controlled servomechanisms*, Zh. tech. fiz., Vol. VIII, No. 10 (1938).
91. E. G. Uderman, *An account of friction and zones of insensitivity of a linear servomechanism with a constant velocity input*, Vestnik elektropromyshlennosti, Nos. 10–11 (1945).
92. V. B. Umanskii, *The effect of friction in a valve on the process of indirect control*, Trudy Dnepropetrovskogo gornogo instituta, Nos. 1–2 (1940).
93. A. A. Fel'dbaum, *Introduction to the Theory of Non-linear Circuits* (Vvedeniye v teoriyu nelineinykh tsepei), Moscow, Gosenergoizdat (1948).
94. A. A. Fel'dbaum, *On the synthesis of optimal systems with the aid of phase spaces*, Avtomat. i Telemekh., No. 2 (1955).
95. A. A. Fel'dbaum, *Optimal processes in systems of automatic control*, Avtomat. i Telemekh., No. 6 (1953).
96. A. A. Fel'dbaum, *The simplest relay systems of automatic control*, Avtomat. i Telemekh., No. 4 (1949).
97. Ya. Z. Tsypkin, *Forced oscillations in relay systems of automatic control*, Avtomat. i Telemekh., No. 5 (1952).
98. Ya. Z. Tsypkin, *On the stability of the periodic states in relay systems of automatic control*, Avtomat. i Telemekh., No. 5 (1953).
99. Ya. Z. Tsypkin, *Transient and Steady Processes in Impulse Circuits* (Perekhodnyye i ustanovivshiyesya protsessy v impul'snykh tsepyakh), Moscow, Gosenergoizdat (1951).
100. Ya. Z. Tsypkin, *The stability and auto-oscillations of relay systems of automatic control*, Proceedings of the Leningrad Red-Flag M.A.E.A., Issue 32 (1950).

101. YA. Z. TSYPKIN, *The frequency method of investigation of periodic states in relay systems of automatic control,* Collection compiled in memory of Acad. A. A. Andronov, Izd. akad. nauk SSSR (1955).
102. YA. Z. TSYPKIN, *The Theory of Relay Systems* (Teoriya relinykh sistem), Moscow, Gostekhizdat (1955).
103. YA. Z. TSYPKIN, *On the connexion between the equivalent coefficient of amplification of a non-linear element and its characteristic,* Avtomat. i Telemekh., No. 4 (1956).
104. M. A. AIZERMAN and F. R. GANTMAKHER, *The determination of periodic states in systems with piecewise-linear characteristics, formed from stages parallel to two given straight lines,* Avtomat. i Telemekh., Nos. 1 and 2 (1957).
105. M. A. AIZERMAN and F. R. GANTMAKHER, *On a class of dynamic problem, connected with the theory of relay systems,* Prikl. mat. i mekh., Vol. XIX, No. 2 (1955).
106. M. A. AIZERMAN and F. R. GANTMAKHER, *A first approximation to the discontinuous systems,* Prikl. mat. i mekh., Vol. XXI, No. 5 (1957).
107. M. A. AIZERMAN and F. R. GANTMAKHER, *On some peculiarities of switchings in non-linear systems of automatic control with piecewise-linear characteristics,* Avtomat. i Telemekh., No. 11 (1957).
108. M. A. AIZERMAN and I. M. SMIRNOVA, *On the application of the small parameter method to the investigation of periodic states in systems of automatic control,* in the Collection in memory of Acad. A. A. Andronov, Moscow, Izd. akad. nauk SSSR (1955).
109. M. A. AIZERMAN, *The effect of non-linear characteristics on the convergence of the automatic control process and on the conditions of the generation of oscillations,* Izv. akad. nauk SSSR, Otdel. takh. nauk, No. 12, p. 148 (1944).
110. M. A. AIZERMAN, *On the determination of critical and non-critical sections on the boundary of stability,* Prikl. mat. i mekh., Vol. XIV, Issue 4 (1950).
111. M. A. AIZERMAN, *On the construction of resonance graphs for systems with non-linear feedback,* Inzh. sbornik, Vol. XIII, p. 151 (1952).
112. M. A. AIZERMAN and F. R. GANTMAKHER, *The determination of the periodic states in non-linear systems of automatic control with piecewise-linear characteristics,* Prikl. mat. i mekh., Vol. XX, No. 1 (1956).
113. M. A. AIZERMAN, *Dry Friction in Problems of Control* (Lectures to the seminar on the theory of automatic control) (Sukhoye treniye v zadachakh regulirovaniya) (Lektsii na seminarye poeorii avtomaticheskogo regulirovaniya), Moscow, Mashgiz (1950).
114. M. A. AIZERMAN, *The physical basis for the application of small parameter methods to the solution of non-linear problems of the theory of automatic control,* Automat. i Telemekh., No. 5 (1953).
115. M. A. AIZERMAN and I. M. SMIRNOVA, *Remarks on the article of E. P. Popov,* Izv. akad. nauk SSSR, Otdel. tekh. nauk, No. 9 (1954).
116. M. A. AIZERMAN and I. M. SMIRNOVA, *On the application of small parameter methods to the investigation of the periodic states in systems of automatic*

control which do not contain a small parameter, in the Collection in memory of Acad. A. A. Andronov, Izd. akad. nauk SSSR (1954).

117. J. R. Dutilh, *Théorie des servomechanismes à relais, L'Onde Électrique*, Vol. 30, p. 438 (1950).
118. H. L. Hazen, *Theory of Servomechanisms*, Journ. Frankl. Inst., Vol. 218, p. 279 (1934).
119. E. C. Johnson, *Sinusoidal Analysis of Feedback Control Systems Containing Non-linear Elements*, Trans. AIEE, Vol. 71, p. 169 (1952).
120. R. Kochenburger, *Frequency Response Method for Analyising and Synthesising Contactor Servomechanisms*, Trans. AIEE, Vol. 69 (1950).
121. R. Kochenburger, *Analysing Contactor Servomechanisms by Frequency-Response Methods*, Electr. Engin. VIII. (1950).
122. H. Léauté, *Du mouvement troublé des moteurs consécutifs à une perturbation brusque, Nouvelle méthode graphique pour l'étude complète de ce mouvement*, Journ. de l'École Polytechnique, Paris No. 61, p. 1 (1891).
123. H. Léauté, *Mémoires sur les oscillations à longues périodes dans les machines actionnées par des moteurs hydrauliques*, Paris, Journ. de l'École Polytechnique, Vol. LV, p. 1 (1895).
124. L. F. A. Lecornu, *Régularisation du mouvement dans les machines*, Gauthier—Villars, Paris, p. 218 (1898).
125. Ph. D. Minorsky, *Introduction to Non-linear Mechanics*, Edwards Ann. Ar bor., (1917).
126. G. Mises, *Zur Theorie des Regulatoren*, Elektrotechnik und Maschinenbau, No. 37, p. 783 (1908).
127. R. Mises, *Dynamische Probleme der Maschinenlehre*, Enc. der mathematischen Wissenschaften, Vol. IV, Issue 2, p. 254 (1911).
128. R. Proell, *Über den indirekt wirkenden Regulierapparat "Patent Proell"*, Z. VDL, Vol. XXVIII, No. 24 (1884), Vol. XXVII, No 25 (1884).
129. J. Rocard, *Dynamique Générale des Vibrations*, Masson (1949).
130. J. Rocard, *Théorie des Oscillateurs.*
131. Lefschetz Solomon, *Contributions to the Theory of Non-linear Oscillations*, Princeton University Press (1950).
132. H. Weiss, *Analysis of relay servomechanisms*, Journ. of Aeronautical Sciences, Vol. 3, No. 7 (1946).
133. K. W. Wagner, *Einführung in die Lehre von den Schwingungen und Wellen* Dietrichsche Verlagsbuchhandlung, Inh. W. Klein, Wiesbaden (1947),

Appendix

1. B. van der Pol and H. Bremmer, *Operational Calculus Based on the Two-Sided Laplace Integral* (Operatsionnoye ischisleniye na osnovye dvustoronnyego preobrazovaniya) Moscow, Inostran. lit. (1953).
2. M. Gardner and J. Barnes, *Transient Processes in Linear Systems with Lumped Constants* (Perekhodnye protsessy v lineinykh sistemakh s sosredotochennymi postoyannymi), translated from the English, Izd. 2-e sipr, Moscow, Gostekhizdat (1951).

3. H. S. Carslaw and J. C. Jaeger, *Operational Methods in Applied Mathematics* (Operatsionnyye metody v prikladnoi matematike), Moscow, Inostran. lit. (1948).
4. D. R. Carson, *Electrical non-stationary phenomena and operational calculus* (Elektricheskiye nestatsioniarnye yavleniya i operatsionnoye ischisleniye), Kharkov—Kiev, DNTVU (1934).
5. M. I. Kontorovich, *Operational Calculus and Non-stationary Phenomena in Electrical Circuits* (Operatsionnoye ischisleniye i nestatsionarnye yavleniya v elektricheshikh tsepyakh), Moscow, Gostekhizdat (1955).
6. A. I. Lur'ye, *Operational Calculus and Its Application to Problems of Mechanics* (Operatsionnoye ischisleniye i yego prilozheniya k zadacham mekhaniki), Moscow, Gostekhizdat (1950).
7. V. V. Solodovnikov, *Introduction to Statistical Dynamics of a System of Automatic Control* (Vvedeniye v statisticheskuyu dinamiku sistemy avtomaticheskogo regulirovaniya), Moscow, Gostekhizdat (1952).
8. A. A. Kharkevich, *Spectra and their Analysis* (Spektry i analiz), Moscow, Gostekhizdat (1952).
9. K. W. Wagner, *Operatorenrechnung nebst Anwendungen in Physik und Technik*, Leipzig (1940).

SUBJECT INDEX